



King's Research Portal

DOI:

[10.1038/s41564-017-0063-9](https://doi.org/10.1038/s41564-017-0063-9)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Pollpeter, D., Parsons, M., Sobala, A., Coxhead, S., Lang, R. D., Bruns, A. M., Papaioannou, S., McDonnell, J. M., Apolonia, L., Chowdhury, J. A., Horvath, C. M., & Malim, M. (2018). Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted antiviral functions of APOBEC3G. *Nature Microbiology*, 3(2), 220-233. <https://doi.org/10.1038/s41564-017-0063-9>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted anti-viral functions of APOBEC3G

Darja Pollpeter¹, Maddy Parsons², Andrew E. Sobala¹, Sashika Coxhead¹,
Rupert D. Lang¹, Annie M. Bruns³, Stelios Papaioannou¹, James M.
McDonnell², Luis Apolonia¹, Jamil A. Chowdhury¹, Curt M. Horvath³, Michael
H. Malim^{1*}

1- Department of Infectious Diseases, King's College London, London SE1 9RT, UK

2- Randall Division of Cell and Molecular Biophysics, King's College London, London SE1 1UL, UK

3- Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

*Corresponding Author: Michael H. Malim
michael.malim@kcl.ac.uk
ORCID iD: 0000-0002-7699-2064

Abstract

Following cell entry, the RNA genome of HIV-1 is reverse transcribed into double-stranded DNA that ultimately integrates into the host cell genome to establish the provirus. These early phases of infection are notably vulnerable to suppression by a collection of cellular anti-viral effectors, called restriction or resistance factors. The host anti-viral protein APOBEC3G (A3G) antagonizes the early steps of HIV-1 infection through the combined effects of inhibiting viral cDNA production, and cytidine-to-uridine driven hypermutation of this cDNA. In seeking to address the underlying molecular mechanism for inhibited cDNA synthesis, we developed a deep sequencing strategy to characterize nascent reverse transcription products and their precise 3'-termini in HIV-1 infected T cells. Our results demonstrate site- and sequence-independent interference with reverse transcription, which requires A3G's specific interaction with reverse transcriptase (RT) itself. This approach also

38 established, contrary to current ideas, that cellular uracil base excision repair
39 (UBER) enzymes target and cleave A3G-edited uridine-containing viral cDNA.
40 Together, these findings yield further insights into the regulatory interplay
41 between RT, A3G and the cellular DNA repair machinery, and identify the
42 suppression of HIV-1 RT by a directly interacting host protein as a new cell-
43 mediated anti-viral mechanism.
44

45 **Introduction**

46

47 Prominent among the anti-retroviral restriction factors^{1,2} are the APOBEC3
48 (A3) proteins, a family of cytidine deaminases with DNA editing capability^{3,4}.
49 Humans encode seven A3 proteins of which APOBEC3G (A3G), A3F, A3D
50 and some allelic forms of A3H are packaged into HIV-1 virions and inhibit
51 infection⁵⁻¹⁰. During wild type HIV-1 infection, A3G and other A3 proteins are
52 largely disabled by the virus-encoded Vif protein, which promotes
53 proteasomal degradation through recruitment to a CUL5 ubiquitin ligase
54 complex¹¹⁻¹³. Vif is therefore essential for productive HIV-1 replication in its
55 natural cell targets, CD4-positive T-cells¹⁴.

56

57 A3 proteins dramatically alter the nascent DNA products of reverse
58 transcription in ensuing rounds of infection: these effects are most
59 pronounced for A3G, which displays the strongest anti-viral phenotype^{6,15}. Not
60 only is the viral complementary DNA (cDNA) littered with numerous cytidine-
61 to-uridine mutations¹⁶⁻¹⁸, but the accumulation of cDNA is also severely
62 reduced compared to infection in the absence of A3 proteins^{14,17,19-22}.

63

64 Figure 1a illustrates three potential, non-mutually exclusive mechanisms for
65 A3G anti-viral action. 1) Hypermutation: the post-synthetic deamination of
66 cytidine to uridine in minus strand reverse transcripts registers as plus strand
67 guanosine-to-adenosine mutations. At least 10% of guanosines can be
68 mutated¹⁴, which severely compromises viral sequence integrity and is
69 typically lethal. 2) cDNA fragmentation and degradation: nascent cDNA is
70 hypermutated as in the first model but with the uracils subsequently
71 recognized by uracil-DNA glycosylases (UDGs) in the uracil base excision
72 repair (UBER) pathway, leading to base excision, recognition by abasic site
73 endonucleases (APE1 or APE2), DNA cleavage and, potentially, degradation.
74 3) Deamination-independent inhibition of reverse transcription: A3G interferes
75 with the biosynthetic capability of the reverse transcriptase (RT) enzyme,
76 reducing the production of cDNA.

77

78 While A3-mediated hypermutation is well documented^{3,4}, the two latter
79 mechanisms for reducing cDNA accumulation remain controversial and poorly
80 understood. Whether UDG-dependent processing of A3G-edited viral DNA
81 occurs is unresolved. Although reducing levels of the UDG, UNG2, has been
82 shown to diminish A3G anti-viral activity^{23,24}, other groups have found that
83 UDGs are dispensable for A3G function and the reduction of cDNA levels²⁵⁻²⁷.
84 A3G's deaminase-independent activity was originally noted when
85 enzymatically-deficient A3G mutant proteins were shown to exhibit a degree
86 of anti-viral function and to suppress cDNA production^{21,28}.

87

88 A common observation from studies of the deaminase-independent effects of
89 A3G is the increasing magnitude of inhibition over the course of reverse
90 transcription and integration, which implies features of processivity. In addition
91 to proposals that A3G inhibits specific steps, such as DNA strand transfers
92 and integration^{27,29}, we have therefore suggested more general interference
93 with RT-catalyzed nucleic acid polymerization^{14,20}. Conceptually, this could
94 occur via a "roadblock" mechanism, in which A3G binds the template genomic
95 RNA and sterically hinders RT translocation, or via direct interference with the
96 biosynthetic capability of RT (or a combination of both). The first hypothesis is
97 supported experimentally by *in vitro* reconstituted primer extension assays
98 where A3G induces pausing/arrest at specific sites in the U5-R sequence²²
99 and by nucleic acid binding studies using purified A3G³⁰, whereas A3G's
100 interaction with RT may lend support to the second³¹.

101

102

103 **Results**

104

105 **High throughput sequencing strategy for determining 3'-termini of HIV-1** 106 **reverse transcripts**

107 We reasoned that defining populations of nascent viral cDNAs produced
108 during HIV-1 infection would shed light upon the mechanism(s) of A3G-
109 mediated viral inhibition. Here, we developed an Illumina MiSeq™ based
110 deep sequencing method that enables unbiased mapping of the 3'-termini of
111 reverse transcription products at single nucleotide resolution in infected T
112 cells. This protocol focuses on early cDNA species, namely the minus strand
113 strong stop ((-)sss) intermediate (comprising the U5 and R regions of the 5'-
114 long terminal repeat) and early first strand transfer products^{32,33}(Figure 1b).
115 The strategy for library creation is depicted in Figure 1c (Supplementary
116 Figure 1 and online methods). Briefly, T cells are infected with HIV-1, nucleic
117 acids extracted and then enriched for viral DNA using capture
118 oligonucleotides. The open 3'-termini of single stranded cDNA are ligated to a
119 single stranded DNA adaptor (Supplementary Figure 1c), and PCR-amplified
120 for sequencing. Barcoding is used to discard non-unique reads.

121

122 **Profiles of cDNA intermediates in T cells infected with HIV-1ΔVif in the** 123 **presence or absence of A3G**

124 To investigate A3G's effects on reverse transcription, HIV-1ΔVif virions were
125 produced without, or with low or high amounts of packaged A3G (Figure 1d).
126 Both infectivity, measured by challenging TZM-bl reporter cells (Figure 1e),
127 and cDNA accumulation in CEM-SS cells at 4 h post infection, measured by
128 qPCR of total DNA extracts (Figure 1f), inversely correlated with the level of
129 packaged A3G. The suppression of cDNA production was characteristically
130 less pronounced than the inhibition of infectivity^{20,29,34}, presumably because
131 the latter reflects the cumulative effects of deaminase-dependent and -
132 independent activities.

133

134 The DNA extracts were then used to prepare libraries for sequencing (Figure
135 1c). We first analyzed cDNA 3'-termini: unique reads for each (-)sss cDNA 3'-

136 end were counted and divided by total read number, which ranged between
137 26500 (high A3G) and 81000 (no A3G) (Supplementary Figure 4b), to obtain
138 relative distributions of cDNA lengths. The top panel in Figure 1g depicts the
139 profile for HIV-1ΔVif without A3G, with the most abundant species being the
140 180 nt (-)sss cDNA (>17% in this experiment). The rest of the profile was
141 notably flat and evenly distributed, with some uplift in the abundance of
142 shorter cDNAs. This pattern differs substantially from what is seen in
143 reconstituted primer extension assays where DNA synthesis pauses at
144 specific sites (Supplementary Figure 3a)³⁵⁻³⁷.

145
146 The presence of A3G resulted in the appearance of five prominent peaks,
147 each representing a specific 3'-terminus (Figure 1g, middle and bottom
148 panels). Importantly, these sites do not match A3G-induced pause sites seen
149 in reconstituted reactions containing purified A3G (Supplementary Figure 3a,
150 lanes 6 to 9)²². Instead, these sites lie one nt 5' to cytidine-to-uridine
151 mutations identified in longer DNA reads (dashed red lines), and occurring at
152 consensus A3G editing sites^{5,16,38,39}. These peaks also featured in first strand
153 transfer sequences, but were absent when catalytically inactive A3G mutant
154 proteins were tested (Supplementary Figures 4a and 5)²⁸. These observations
155 suggested that A3G editing hotspots serve as cDNA cleavage sites.

156

157 **Target cell UBER enzymes mediate the detection and cleavage of A3G-** 158 **edited HIV-1 cDNA**

159 To test whether UDGs, specifically UNG2^{40,41}, mediate the processing of
160 uridine-containing cDNA, leading to endonucleolytic cleavage, we utilized the
161 bacteriophage uracil-DNA glycosylase inhibitor (UGI)⁴². CEM-SS target cells
162 and HEK293T producer cells stably expressing codon optimized UGI (hUGI)²⁵
163 were generated, the latter being required because UNG2 is packaged into
164 virions^{43,44}. Enzymatic analyses of cell lysates confirmed efficient UDG
165 suppression (Supplementary Figure 6). HIV-1ΔVif was produced in control or
166 hUGI expressing HEK293T cells in the absence or presence of A3G (Figure
167 2a). In bulk measurements of virus infectivity and cDNA abundance (Figures
168 2b and 2c), no differences were attributable to hUGI, irrespective of

169 expression in producer, target or both cultures, reconfirming that UNG2
170 inhibition does not impact A3G anti-viral function²⁵⁻²⁷. Sequencing analysis
171 revealed that UNG2 inhibition in target cells mitigated A3G's induction of
172 foreshortened cDNAs, with the overall profiles of 3'-termini resembling those
173 seen without A3G (Figure 2d; and Supplementary Figure 7).

174

175 We therefore conclude: first, that target cell UNG2 can detect A3G-edited
176 cDNAs leading to uracil removal and subsequent cleavage (Figure 1a,
177 pathway 2); second, that UBER-mediated cDNA fragmentation does not lead
178 to their complete degradation, implying that a deamination-independent
179 mechanism likely underlies the inhibition of cDNA accumulation by A3G; and,
180 third, that A3G does not induce site-specific RT pausing, thus arguing against
181 the aforementioned "roadblock" mechanism for RT inhibition.

182

183 **Interaction between A3G and the HIV-1 RT heterodimer**

184 The 3'-termini profiles seen in the absence of A3G, in its presence but with
185 UNG2 activity abrogated, or with catalytically inactive A3G mutants are very
186 similar to each other, yet levels of cDNA production differ substantially. One
187 model to reconcile these observations is for A3G to affect RT processivity
188 through a mechanism that lacks template sequence specificity. This would
189 yield similarly reduced nucleotide addition efficiency at all positions,
190 regardless of RT location on the template RNA. We hypothesized that this
191 could be achieved through A3G interacting with RT and inhibiting its
192 enzymatic capability.

193

194 To evaluate this model, we initially used lysates from transfected HEK293T
195 cells to confirm that A3G co-immunoprecipitates with the HIV-1 RT
196 heterodimer (p51 and p66 subunits) (Figure 3a)³¹. Since A3G and RT are
197 RNA binding proteins, we employed several techniques to ensure that this
198 interaction is not simply bridged by RNA. First, co-immunoprecipitation assays
199 were performed in the presence of ribonucleases (Figure 3b); although either
200 a heterogeneous RNase mixture (upper panel) or RNase A (lower panel)

noticeably reduced the amount of recovered A3G, the effect was saturable with A3G pull-down still remaining clear at high RNase levels.

We next carried out surface plasmon resonance (SPR) assays using purified p51 and A3G, both stripped of nucleic acids (Supplementary Figure 8). Measurements provided association and dissociation curves that gave acceptable fits to a single interaction model (Figure 3c), suggesting that the majority of complexes were homogeneous in nature. The association rate constant (k_{on}) was calculated to $1.4 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, and the dissociation rate constant (k_{off}) to $2.2 \times 10^{-3} \text{ s}^{-1}$, yielding a dissociation constant (K_d) of $\sim 1.6 \text{ }\mu\text{M}$; a value close to that calculated using rotational anisotropy ($2.3 \text{ }\mu\text{M}$)⁴⁵. Curve fitting indicated a maximal binding (B_{max}) of ~ 200 RU of p51. Since the A3G ligand surface density is ~ 200 RU this is consistent with a 1:1 binding stoichiometry for the A3G-p51 complex. Taken together, these findings indicate that A3G and RT interact directly and that RNA may help to stabilize the complex.

Using cell-based FRET assays to measure the A3G-RT interaction

To address further the direct or indirect nature of the A3G-RT interaction, we turned to fluorescence resonance energy transfer (FRET) assays as they circumvent any need for protein purification or enrichment. Specifically, we employed green fluorescent protein (GFP, the donor) and mCherry (the acceptor) fusion proteins, co-expressed in HeLa cells, together with fluorescence lifetime imaging microscopy (FLIM) for the FRET readout. Here, the average lifetime (τ) of the donor is reduced when in a FRET state owing to the proximity of the acceptor (online methods)⁴⁶. FRET is shown both as representative pseudocolored images with blue/green indicating longer τ than yellow/red colors (Figure 3d and e), and as efficiency percentages (Figure 3f).

Cells were transfected with vector pairs expressing the indicated fusion proteins, fixed and analyzed after 24 h. As expected, A3G_GFP localized to P-bodies, stress granules as well as the cytosol, while the RT subunits were more evenly distributed across the cytoplasm (Figure 3d and e). The oligomerization of A3G⁴⁷⁻⁴⁹ and hetero-dimerization of RT subunits provided

235 positive controls, each visualized as yellow/orange coloration (Figure 3e, top
236 row; and Figure 3f). The RNA helicase MOV10, which co-localizes to P-
237 bodies with A3G, but is not thought to interact with A3G directly since co-
238 immunoprecipitation depends on RNA bridging⁵⁰ served as a negative control.
239 No change in donor τ was observed, consistent with the lack of a direct A3G-
240 MOV10 interaction. The low FRET efficiency displayed between A3G_GFP
241 and the oligomerization-deficient W127A_mCherry mutant⁵¹ provided further
242 assay validation (Figure 3e, second row). In clear contrast, co-expression of
243 A3G_GFP with p51_mCherry or p66_mCherry yielded marked reductions in τ ,
244 demonstrating very close proximity between RT subunits and A3G (Figure 3e,
245 third row), consistent with a direct protein-protein interaction.

246

247 **A3G and RT interact within intact HIV-1 virions**

248 We also adapted our FRET-FLIM system for cell-free, bulk HIV-1 particles
249 (Figure 4). A3G_GFP was packaged into virions by co-expression with HIV-
250 1 Δ Vif, and these were purified and immobilized on coverslips, before immuno-
251 staining and FRET analysis. Rather than using protein fusions that could
252 interfere with viral assembly or particle, RT was detected using a Cy3-labelled
253 anti-RT Fab fragment. A3G_GFP exhibits a normal GFP lifetime in the
254 absence of a fluorescent acceptor (Figure 4a) and a positive control of co-
255 packaged A3G_GFP and A3G_mCherry demonstrates the suitability of this
256 assay for detecting protein-protein interactions within virions (Figure 4b, left
257 panel, top row; and Figure 4c). A substantial increase in FRET efficiency was
258 evident when RT was detected with labeled Fab (Figure 4b, right panel,
259 bottom row; and Figure 4c), which was not seen with other packaged proteins
260 as acceptors (Vpr_GFP or cyclophilin A_GFP), the antibody-mediated
261 detection of Capsid protein (p24^{Gag}) (Figure 4b, top right panels; and Figure
262 4c), or for virions engineered to lack RT (Figure 4b, left panel, bottom row).
263 These data demonstrate that A3G and RT interact in viral particles, and
264 support the view that A3G is positioned at the site of reverse transcription.

265

266 **Regions of A3G that interact with HIV-1 RT**

Regions of A3G that interact with RT were mapped by co-expression of RT with a set of A3G-glutathione-S-transferase (GST) fusions and co-immunoprecipitation (Figure 5a). Consistent with earlier results³¹, full length A3G and a fragment spanning residues 65 to 132 interacted with RT (lanes 3 and 4). Interestingly, the N-terminal 76 amino acid fragment of A3G also bound RT (lane 5), with further truncations revealing dependency on the residues between positions 30 and 42 (lanes 7 and 8). Thus, RT interacts with two discrete and non-overlapping regions of A3G.

Double alanine scanning mutagenesis between residues 30 and 42 in the context of full length A3G showed that a W34A/L35A double mutant, and subsequently the single L35A mutant, were impaired for interacting with RT (Figure 5b). These mapping data were confirmed in intact cells, as the substitution of leucine at position 35 with alanine (W34A/L35A and L35A mutants) significantly reduced FRET efficiency (Figure 5c, row 3 and 4 and Figure 5d). As a control, the R24A mutation in A3G, which disrupts a positively charged patch at the A3G dimer interface⁵² and inhibits RNA binding⁵¹ (Supplementary Figure 3c) was shown not to alter binding to RT (Figure 5b, c and d), thereby reinforcing the conclusion that the A3G-RT interaction is RNA independent. Importantly, like R24A⁵¹, the L35A protein maintains full enzymatic activity as measured in an *E. coli*-based DNA editing assay (Supplementary Figure 9), indicating that these substitutions do not cause global structural defects or misfolding.

Inhibition of the RT interaction diminishes A3G anti-viral activity

Derivation of a mutation in A3G that interferes with RT binding but maintains deaminase function allowed us to evaluate directly the importance of the A3G-RT interaction for anti-viral activity. The L35A protein, similar to R24A⁵¹, is not packaged into viruses as efficiently as wild type A3G; however, this defect is partial and is overcome by increasing protein expression in virus producer cells.

For these experiments, we used titrations of wild type A3G in producer cells to yield viruses with a range of packaged A3G levels (Figure 6a; Supplementary

301 Figure 10a). These stocks were then used to infect TZM-bl cells to determine
302 overall infectivity (Supplementary Figure 10b), and CEM-SS cells to measure
303 cDNA production (Supplementary Figure 10c); fitted curves were generated
304 that related infectivity or cDNA levels to amounts of packaged A3G
305 (Supplementary Figure 10d and 10e).

306

307 Viruses carrying the L35A or R24A mutants were prepared in parallel and
308 analyzed for packaging, infectivity and cDNA production. By extrapolation
309 along the fitted curves, we compared their effects on infectivity (Figure 6b)
310 and reverse transcription (Figure 6c) with those of wild type A3G at matched
311 packaging levels. Over eight experiments, L35A was significantly less anti-
312 viral than A3G, and this was coupled with increased cDNA production.
313 Critically, R24A displayed equivalent anti-viral activity to wild type A3G at
314 comparable levels of encapsidation, thus phenotypically distinguishing itself
315 from L35A. Moreover, the finding that R24A still suppresses reverse
316 transcription (Figure 6c), while binding RNA less well (Supplementary Figure
317 3)⁵¹, is again consistent with RT suppression not being mediated by the
318 “roadblock” mechanism.

319

320 To test whether the remaining anti-viral activity of L35A is deaminase
321 dependent, we combined this change with a mutation, C288S, that ablates
322 deamination but incompletely prevents anti-viral activity (Figure 6d; and
323 Supplementary Figures 5, 10g and 10h)^{27,28}. Importantly, the double mutant
324 (C288S/L35A) displayed a total loss of anti-viral function even at high
325 expression levels (Figure 6d), underscoring the importance of both attributes
326 for full A3G activity.

327

328 Lastly, we analyzed the cDNA profiles in CEM-SS cells challenged with HIV-
329 1ΔVif lacking A3G or containing similar levels of wild type A3G, R24A or L35A
330 (Figure 6a and e). The three profiles obtained with A3G-bearing viruses were
331 essentially identical, confirming the editing competency of these mutant
332 proteins and the absence of any site-specific pausing that correlated with
333 inhibited cDNA synthesis. We therefore conclude that A3G regulates HIV-1

334 RT through a direct interaction that impedes enzymatic capability and
335 suppresses the production of DNA evenly and without sequence or site
336 specificity (Figure 1a, pathway 3).
337

Discussion

Here, we present a custom method for monitoring the progression of HIV-1 reverse transcription at single nucleotide resolution in infected cells. Our findings have relevance for understanding not only A3-mediated restriction, but also fundamental aspects of reverse transcription. In particular, much of the current knowledge of reverse transcription has been obtained using *in vitro* reconstituted systems^{32,33,53,54} (e.g., Supplementary Figure 3). In agreement with such studies, we demonstrate a marked accumulation of a predominant (-)sss species (e.g., Figure 1g)^{35,55,56}, yet defining its precise 3'-terminus has previously been elusive. The abundant (-)sss cDNA we detect has a single 3'-cytidine (nt position 180), and we presume this to be derived from the ^{Cap}G1-form of HIV-1 RNA. This is consistent with the preferential dimerization and packaging of ^{Cap}G1 RNA that has a single guanosine at its 5'-terminus, rather than the ^{Cap}G2- or ^{Cap}G3-forms that contain two or three 5'-guanosines^{57,58}.

In contrast to the distinctive (-)sss product seen in living cells and reconstituted systems, the foreshortened cDNAs that are detected in reconstitution experiments are negligible in infected cells (Figure 1g; Supplementary Figure 3). Their formation has been ascribed to RT pausing/arrest at homopolymeric motifs or RNA secondary structures^{35-37,59}. These reproducible disparities may reflect: variations in the enzymatic capabilities of HIV-1 RT in the microenvironment of reverse transcription complexes in infected cells versus in reconstituted reactions; differences in the formation of viral genomic RNA secondary and tertiary structures; or the presence or relative abundance of viral (or host) proteins such as the HIV-1 Nucleocapsid protein at the site of reverse transcription.

Not surprisingly, experimental system dependent differences extend to the effects of A3G on reverse transcription. Studies with reconstituted assays had indicated A3G-induced pausing of RT at specific sites²² (Supplementary Figure 3a), which helped formulate the "roadblock" model. Critically, our cell-

371 based DNA sequencing approach provided no evidence for localized A3G-
372 induced RT pausing, with the prominent A3G-dependent cDNA peaks being
373 caused by UBER activity (Figure 2). These findings evoke a sequence- and
374 site-independent mechanism for the suppression of RT by A3G, and imply:
375 first, that there are differences in RT's behavior in the context of HIV-1
376 infection; and, second, that the A3G-induced pausing that is seen in
377 reconstituted reactions reflects an assay-specific epiphenomenon.

378

379 Intrinsic to our model for A3G function is its negative regulatory interaction
380 with RT, which, based on a series of corroborating biochemical, biophysical
381 and cell-based FRET assays (Figures 3 to 5)³¹, we consider to be dependent
382 on direct binding. Purposeful interference with this interaction through
383 replacement of leucine-35 resulted in a significant loss of anti-viral activity
384 (Figure 6b and c), with the remaining anti-viral activity of L35A being
385 attributable to its deaminase function since the C288S/L35A double mutant
386 displayed a complete loss of anti-viral phenotype (Figure 6d).

387

388 Our sequencing approach also allowed us to examine the effects of UBER on
389 viral cDNA following exposure to A3G. Although, we reveal UNG2-mediated
390 uridine excision from deaminated cDNA, our measurements show ensuing
391 endonucleolytic cleavage is far from complete since substantial proportions of
392 cytidine-to-uridine edited cDNAs remained intact. (Figures 1, 2 and 6).
393 Appreciating this inefficiency helps to reconcile the apparent discrepancy
394 between the predicted degradative fate of deaminated cDNAs (Figure 1a,
395 mechanism 2) and the lack of persuasive evidence for UDG involvement in
396 A3G function. Nevertheless, despite the absence of an evident anti-viral effect
397 in single round infections (Figure 2b), inefficient UBER recognition of edited
398 cDNA could still play a role in the interplay between HIV-1 and infected cells,
399 for instance through sensing of aberrant cDNA fragments as a pathogen-
400 associated molecular pattern^{1,60}. How HIV-1 may circumvent this (aside from
401 Vif induced A3 destruction in virus producing cells) is a matter of conjecture,
402 but we note that HIV-1 Vpr and its interacting CUL4 ubiquitin ligase induces
403 UNG2 degradation^{23,44,61}.

404

405 Uracilation of HIV-1 cDNA can also arise from the misincorporation of dUTP in
406 place of dTTP during reverse transcription⁶², especially in non-dividing cells
407 such as macrophages that have high dUTP:dTTP ratios^{63,64}. UNG2-mediated
408 recognition of the resulting uracilated viral cDNA suppresses infection, with
409 restriction being most evident at the level of DNA integration⁶⁵⁻⁶⁸. Interestingly,
410 and in contrast to observations made with A3G-induced uracilation (Figure 2,
411 Supplementary Figure 7)²⁵⁻²⁷, UGI-mediated suppression of UNG2 in the face
412 of misincorporation-driven uracilation provokes viral rescue^{65,66}. The basis for
413 this dichotomy remains to be determined, but could relate to variations in the
414 extent of uracilation, or cell type dependent differences in the efficiency of
415 UBER-mediated cDNA recognition or the fate/processing of uracilated cDNA.

416

417 Reverse transcription is a vulnerable step in the HIV-1 life cycle: it is highly
418 susceptible to pharmacological inhibition, and is also suppressed by A3G and
419 other host cell restriction factors. TRIM5 α destabilizes post-entry viral capsids
420 which compromises the timely onset/completion of reverse transcription⁶⁹⁻⁷¹,
421 while the deoxynucleotide triphosphohydrolase SAMHD1 interferes with
422 reverse transcription by depleting cellular dNTP pools². In contrast to these
423 indirect mechanisms, we demonstrate that RT is inhibited by A3G by a distinct
424 mechanism, namely via a direct protein interaction that impedes biosynthetic
425 activity. Establishing the concept that HIV-1 RT can be negatively regulated in
426 this manner raises the possibility that additional cellular factors may modulate
427 RT⁷².

428

429 Future directions for this work include determining: how A3G binding to RT
430 interferes with enzymatic function, how leucine-35 (a residue that is highly
431 conserved within the A3 family) participates in the RT interaction, and which
432 regions of RT bind to A3G. Elucidating the macromolecular structure of the
433 A3G-RT complex will naturally be a major advance, and can assist in
434 resolving the heretofore elusive structure of full-length A3G⁷³. Finally, by
435 revealing that a non-virally encoded interacting partner can regulate HIV-1
436 RT, we highlight a previously unrecognized vulnerability in the enzyme that
437 may have the potential to be targeted therapeutically.

438 **Acknowledgments:**

439 We would like to acknowledge the support and fruitful debate with members of
440 the Malim laboratory, the insights of Jernej Ule and Rebecca Oakey on
441 optimization of the sequencing protocol, and Michael Emerman and Stephen
442 Hughes for the generous provision of reagents. We also thank Matt Arno at
443 the King's College London Genomic Center and Debbie Hughes at the
444 University College London (UCL) Institute for Neurology Next Generation
445 Sequencing Facility for help with MiSeq sequencing runs. The work was
446 supported by the U.K. Medical Research Council (G1000196 and
447 MR/M001199/1 to M.M.; and MR/K015664/1 to M.P.), the Wellcome Trust
448 (106223/Z/14/Z to M.M.), the European Commission's Seventh Framework
449 Programme [FP7/2007-2013] under grant agreement n°PIIF-GA-2012-329679
450 (to D.P.), King's alumni community sponsored King's Undergraduate
451 Research Fellowships (to R.D.L), King's School of Medicine Summer
452 Studentship Award (to J.C), and the Department of Health via a National
453 Institute for Health Research comprehensive Biomedical Research Centre
454 award to Guy's and St. Thomas' NHS Foundation Trust in partnership with
455 King's College London and King's College Hospital NHS Foundation Trust
456 (guysbrc-2012-1).

457

458 **Author Contributions:**

459 D.P. co-wrote the manuscript and executed all experiments with the following
460 exceptions: R.D.L. performed the co-immunoprecipitation shown in Figure 5a,
461 M.P. carried out all the microscopy and FRET-FLIM experiments. A.E.S.
462 wrote and ran the analysis software for analyzing raw FASTQ sequencing
463 data. S.C. carried out the double alanine scan for the A3G-RT binding site
464 mapping. A.M.B. and C.M.H carried out and analyzed the single molecule
465 RNA binding assays (Supplementary Figure 3c). S.P., R.D.L. and J.C.
466 contributed to reagent generation, in particular for Fig 5. J.M.M. contributed to
467 the SPR experiments (Figure 3c) and performed the analysis. L.A and A.E.S
468 contributed to the sequencing library design. D.P and M.H.M. conceived the
469 experiments and co-wrote the manuscript. All authors cross-checked the
470 manuscript.

471

472 **Competing financial interest:**

473 The authors have no conflict of financial interest.

474

475 **Materials and Correspondence:**

476 Correspondence should be send to Michael Malim at

477 michael.malim@kcl.ac.uk

478

479 **Figure Legends**

480

481 **Figure 1: Effects of A3G on profiles of nascent HIV-1 cDNA products in** 482 **infected T cells.**

483 **a)** Early steps of the HIV-1 life cycle illustrating three proposed anti-retroviral
484 mechanisms for A3G that are deaminase-dependent (pathways 1 and 2) or -
485 independent (pathway 3). **b)** Diagram of HIV-1 reverse transcription. The first
486 full intermediate, (-)sss cDNA, is completed in step 3. PBS: primer binding
487 site, PPT: polypurine tract. **c)** Basic steps of sequencing library preparation.
488 During infection, HIV-1 produces nascent viral cDNAs of increasing length
489 (see step 2 in b). Sequencing reads reveal precise 3'-termini at the points of
490 adaptor-viral DNA ligation (red box). **d)** Immunoblot analysis of HIV-1 virion
491 lysates from one of six independent virus preparations. 'Low' or 'High' A3G
492 refers to producer cell transfection ratios of 1:10 or 1:4, respectively (A3G
493 expression plasmid to NL4.3/ Δ Vif). **e)** Single-cycle virion infectivity measured
494 by β -galactosidase activity in challenged TZM-bl reporter cells. **f)** Quantitative
495 PCR measuring cDNA abundance in CEM-SS cells at 4 h post-infection. For
496 e) and f) the individual data points with their mean and standard deviation of
497 eight independent infections from six virus preparations are shown. ***
498 indicates p-value of <0.0001 in an unpaired, two-tailed t-test with Welch's
499 correction performed in GraphPad Prism[®]. **g)** Numbers of unique sequencing
500 reads ending at each nt of the HIV-1_{NL4.3} (-)sss cDNA were divided by the total
501 read number (Supplementary Figure 4b) within each sample to show the
502 relative abundance of cDNAs for each length between nt positions 23 and
503 182. Shown in dashed red lines are the percentages of reads carrying C to
504 T/U mutations at that position (scale on the right y-axis). See Method section
505 for analysis details. One representative experiment out of three independent
506 repeats is shown.

507 **Figure 2: Consequences of UDG inhibition on A3G antiviral phenotype**
508 **and cDNA profiles**

509 a) Immunoblot analysis of HIV-1 virion lysates showing increasing amounts of
510 packaged A3G_HA at constant CA levels for virions produced in the presence
511 or absence of a codon optimized (humanized) uracil-DNA glycosylase
512 inhibitor (hUGI). 'Low' or 'High' A3G refers to a producer cell transfection
513 ratios of 1:10 or 1:1, respectively (A3G expression plasmid to NL4.3/ΔVif).
514 One of three independent sets of virus preparations used for b) and c) is
515 shown. **b)** Virion infectivity was evaluated by challenging TZM-bl cells and
516 measurement of β -galactosidase activity. **c)** The abundance of (-)ssb
517 containing cDNA in CEM-SS cells at 4 h post-infection was measured by
518 quantitative PCR. For b) and c) each viral preparation was used to infect
519 TZM-bl or CEM-SS target cells with or without hUGI, black dots and grey
520 squares respectively. The individual data points with their mean and standard
521 deviation for three independent viral preparations and infections are shown. **d)**
522 Sequencing reads from a MiSeq™ library run were analyzed and presented
523 as in Figure 1g. The labeling to the right indicates whether the HEK293T
524 producer cells (Prod) and/or the CEM-SS target (Target) cells expressed
525 hUGI. No A3G indicates the absence of A3G in producer cells and high A3G
526 refers to relative A3G content in the producer cells. Sequencing data are
527 derived from one representative experiment out of two independent repeats.

528

Figure 3: Interaction of A3G with HIV-1 reverse transcriptase.

Co-immunoprecipitation analysis of A3G_HA binding to FLAG tagged HIV-1 RT. Transfected HEK293T cell lysates were subjected to anti-FLAG immunoprecipitation, recovered proteins were detected with anti-HA (for A3G), anti-RT or anti-FLAG antibodies. CD8_FLAG served as an irrelevant protein control. One representative experiment of three repeats is shown. *HC: immunoglobulin heavy chain **b)** RNase resistance of the A3G-RT complex. Shown are anti-FLAG immunoprecipitations after the bead bound proteins had been subjected to RNase A or RNase Mix treatment, at the indicated concentrations, followed by washing and immunoblotting. One representative experiment of three repeats is shown. Samples without RT_FLAG carry CD8_FLAG as an irrelevant tagged protein control. **c)** Surface plasmon resonance analysis of purified A3G and p51 on a Biacore T-200 instrument. Association and dissociation curves of p51_FLAG to immobilized A3G_6xHis at the indicated concentrations are shown. The sensorgram indicates specific binding between the two components, and the responses gave good fits to a single interaction binding model with a K_d of $\sim 1.6 \mu\text{M}$. **d)-f)** Measurements of FRET efficiency using FLIM in HeLa cells expressing GFP and mCherry fusion proteins. Representative images with GFP fluorescence from multiphoton laser scanning microscopy (left panel) and corresponding wide field CCD camera images of mCherry fluorescence (right panels (e only)) are shown. The center panels represent pseudo-colored images of GFP lifetime (τ) (blue/green, normal/longer GFP lifetime; yellow/red, shorter GFP lifetime indicating FRET). **d)** Control images demonstrating normal GFP lifetime in the absence of mCherry acceptor. White scale bars represent $10 \mu\text{m}$. **e)** Co-expression of indicated GFP and mCherry fusion proteins and the fluorescence lifetime according to the scale in d) indicating the presence or absence of FRET. **f)** Dot plot of FRET efficiencies with their mean and standard deviation from $n=7$ cells each.

560 **Figure 4: A3G interaction with HIV-1 RT in virions**

561 Suspensions of HIV-1 virions with packaged A3G_GFP, GFP_Vpr,
562 GFP_CYP A or A3G_GFP and A3G_mCherry were immobilized on coverslips,
563 fixed and stained with Cy3 labeled anti-RT or anti-CA Fab fragments. **a) and**
564 **b)** Representative images show clusters of HIV-1 virions immobilized on
565 fibronectin streaks with green fluorescence (left panel), red fluorescence (Cy3
566 or mCherry as indicated, right panel) and GFP lifetime as pseudo-colored
567 images according to the indicated scale (as in Figure 3). White scale bars
568 represent 10 μ m. **a)** A3G_GFP demonstrates normal lifetime when packaged
569 into HIV-1 virions. **b)** FRET is detected for the positive control of A3G_GFP
570 and A3G_mCherry (upper left panel) and between A3G_GFP and Cy3 stained
571 RT (lower right panel), but not between Vpr and RT, CYP A and RT, or A3G
572 and CA (upper right panels). The absence of a signal for red fluorescence
573 with HIV-1 Δ RT virions confirmed the specificity of the anti-RT Fab fragments
574 (lower left panel). **c)** Quantification of FRET efficiencies for n=5 areas.
575 Individual measurements with their mean and standard deviation are shown.

576

577

578 **Figure 5: Mapping of A3G-RT interaction sites on A3G protein**
579 **a)** Anti FLAG immunoprecipitation of p51_FLAG and p66_FLAG co-expressed
580 with GST or GST_A3G fusion proteins, recovered proteins were detected with
581 anti-GST (for A3G) or anti-FLAG antibodies as indicated. A3G truncations are
582 indicated and numbers refer to amino acid positions in A3G. **b)** Co-
583 immunoprecipitation analysis of wild type or mutant A3G with HIV-1
584 p51_FLAG and p66_FLAG, recovered proteins were detected with anti-HA
585 (for A3G) or anti-FLAG antibodies. One representative out of three
586 experiments is shown. **c)** FRET-FLIM analysis of wild type or mutant A3G with
587 the p66 subunit of HIV-1 RT. Representative images show green fluorescence
588 (GFP, left panel) and red fluorescence (mCherry, right panel) and GFP
589 lifetime as pseudo-colored images according to the indicated scale (as in Fig
590 5). White scale bars represent 10 μ m. **d)** Dot plots showing individual FRET
591 efficiencies with their mean and one standard deviation from n=12 cells each.
592 *** indicates p-value of <0.0001 in an unpaired, two-tailed t-test performed in
593 GraphPad Prism[®].
594
595

Figure 6: Phenotypes of packaged L35A and R24A A3G mutant proteins on viral infectivity and cDNA profiles

a) Immunoblot analysis of HIV-1 virions showing relative amounts of packaged wild type or mutant A3G_HA at constant CA levels. Ratios refer to the amounts of transfected A3G expression plasmid to proviral plasmid during virus production. **b)** A3G-L35A, but not A3G-R24A, displays diminished HIV-1 inhibitory activity. A3G packaging was quantified by immunoblot density measurements and the different wild type A3G packaging levels were plotted over measured infectivity. The extent of infection inhibition exerted by the wild type protein at the empirically determined level of packaged mutant protein was then extrapolated (see Supplementary Fig 10). Inhibition levels, in % relative to the no A3G control, of wild type A3G (triangles) and L35A or R24A (circles) in eight (L35A) or seven (R24A) independent experiments are shown. A paired, two tailed student t test was performed in GraphPad Prism[®] and * indicates $p < 0.05$ ($p = 0.0223$), ns: not significant **c)** As in b), but with (-)ss cDNA abundance measured by qPCR in cells 4 h post-infection for virions carrying wild type or mutant A3G. ** indicates $p < 0.005$ ($p = 0.0028$). **d)** Relative infectivity of $n = 5$ independent virus preparations carrying the indicated wild type or mutant A3G at equal, 'high' levels as shown in the representative immunoblot in Supplementary Fig 10g. A paired, two tailed student t test was performed in GraphPad Prism[®]. * indicates $p < 0.05$ ($p = 0.0397$ for A3G wt – L35A; $p = 0.0297$ for C288S – C288S/L35A and $p = 0.0137$ for L35A – C288S/L35A). **e)** Sequencing reads from a MiSeq[™] library run were analyzed and presented as in Fig 1g. Labels to the right indicate the presence or absence of A3G proteins in virions.

622 **References:**

623

- 624 1. Doyle, T., Goujon, C. & Malim, M.H. HIV-1 and interferons: who's
625 interfering with whom? *Nat Rev Microbiol* **13**, 403-13 (2015).
- 626 2. Simon, V., Bloch, N. & Landau, N.R. Intrinsic host restrictions to HIV-1
627 and mechanisms of viral escape. *Nat Immunol* **16**, 546-53 (2015).
- 628 3. Desimmie, B.A. et al. Multiple APOBEC3 restriction factors for HIV-1
629 and one Vif to rule them all. *J Mol Biol* **426**, 1220-45 (2014).
- 630 4. Harris, R.S. & Dudley, J.P. APOBECs and virus restriction. *Virology*
631 **479-480**, 131-45 (2015).
- 632 5. Bishop, K.N. et al. Cytidine deamination of retroviral DNA by diverse
633 APOBEC proteins. *Curr Biol* **14**, 1392-6 (2004).
- 634 6. Hultquist, J.F. et al. Human and rhesus APOBEC3D, APOBEC3F,
635 APOBEC3G, and APOBEC3H demonstrate a conserved capacity to
636 restrict Vif-deficient HIV-1. *J Virol* **85**, 11220-34 (2011).
- 637 7. Sheehy, A.M., Gaddis, N.C., Choi, J.D. & Malim, M.H. Isolation of a
638 human gene that inhibits HIV-1 infection and is suppressed by the viral
639 Vif protein. *Nature* **418**, 646-50 (2002).
- 640 8. Apolonia, L. et al. Promiscuous RNA binding ensures effective
641 encapsidation of APOBEC3 proteins by HIV-1. *PLoS Pathog* **11**,
642 e1004609 (2015).
- 643 9. Luo, K. et al. Amino-terminal region of the human immunodeficiency
644 virus type 1 nucleocapsid is required for human APOBEC3G
645 packaging. *J Virol* **78**, 11841-52 (2004).
- 646 10. Soros, V.B., Yonemoto, W. & Greene, W.C. Newly synthesized
647 APOBEC3G is incorporated into HIV virions, inhibited by HIV RNA, and
648 subsequently activated by RNase H. *PLoS Pathog* **3**, e15 (2007).
- 649 11. Sheehy, A.M., Gaddis, N.C. & Malim, M.H. The antiretroviral enzyme
650 APOBEC3G is degraded by the proteasome in response to HIV-1 Vif.
651 *Nat Med* **9**, 1404-7 (2003).
- 652 12. Marin, M., Rose, K.M., Kozak, S.L. & Kabat, D. HIV-1 Vif protein binds
653 the editing enzyme APOBEC3G and induces its degradation. *Nat Med*
654 **9**, 1398-403 (2003).
- 655 13. Yu, X. et al. Induction of APOBEC3G ubiquitination and degradation by
656 an HIV-1 Vif-Cul5-SCF complex. *Science* **302**, 1056-60 (2003).
- 657 14. Gillick, K. et al. Suppression of HIV-1 infection by APOBEC3 proteins
658 in primary human CD4(+) T cells is associated with inhibition of
659 processive reverse transcription as well as excessive cytidine
660 deamination. *J Virol* **87**, 1508-17 (2013).
- 661 15. Phalora, P.K., Sherer, N.M., Wolinsky, S.M., Swanson, C.M. & Malim,
662 M.H. HIV-1 replication and APOBEC3 antiviral activity are not
663 regulated by P bodies. *J Virol* **86**, 11712-24 (2012).
- 664 16. Harris, R.S. et al. DNA deamination mediates innate immunity to
665 retroviral infection. *Cell* **113**, 803-9 (2003).
- 666 17. Mangeat, B. et al. Broad antiretroviral defence by human APOBEC3G
667 through lethal editing of nascent reverse transcripts. *Nature* **424**, 99-
668 103 (2003).

- 669 18. Zhang, H. et al. The cytidine deaminase CEM15 induces
670 hypermutation in newly synthesized HIV-1 DNA. *Nature* **424**, 94-8
671 (2003).
- 672 19. Bishop, K.N., Holmes, R.K. & Malim, M.H. Antiviral potency of
673 APOBEC proteins does not correlate with cytidine deamination. *J Virol*
674 **80**, 8450-8 (2006).
- 675 20. Bishop, K.N., Verma, M., Kim, E.Y., Wolinsky, S.M. & Malim, M.H.
676 APOBEC3G inhibits elongation of HIV-1 reverse transcripts. *PLoS*
677 *Pathog* **4**, e1000231 (2008).
- 678 21. Holmes, R.K., Koning, F.A., Bishop, K.N. & Malim, M.H. APOBEC3F
679 can inhibit the accumulation of HIV-1 reverse transcription products in
680 the absence of hypermutation. Comparisons with APOBEC3G. *J Biol*
681 *Chem* **282**, 2587-95 (2007).
- 682 22. Iwatani, Y. et al. Deaminase-independent inhibition of HIV-1 reverse
683 transcription by APOBEC3G. *Nucleic Acids Res* **35**, 7096-108 (2007).
- 684 23. Schrefelbauer, B., Yu, Q., Zeitlin, S.G. & Landau, N.R. Human
685 immunodeficiency virus type 1 Vpr induces the degradation of the UNG
686 and SMUG uracil-DNA glycosylases. *J Virol* **79**, 10978-87 (2005).
- 687 24. Yang, B., Chen, K., Zhang, C., Huang, S. & Zhang, H. Virion-
688 associated uracil DNA glycosylase-2 and apurinic/aprimidinic
689 endonuclease are involved in the degradation of APOBEC3G-edited
690 nascent HIV-1 DNA. *J Biol Chem* **282**, 11667-75 (2007).
- 691 25. Kaiser, S.M. & Emerman, M. Uracil DNA glycosylase is dispensable for
692 human immunodeficiency virus type 1 replication and does not
693 contribute to the antiviral effects of the cytidine deaminase APOBEC3G.
694 *J Virol* **80**, 875-82 (2006).
- 695 26. Langlois, M.A. & Neuberger, M.S. Human APOBEC3G can restrict
696 retroviral infection in avian cells and acts independently of both UNG
697 and SMUG1. *J Virol* **82**, 4660-4 (2008).
- 698 27. Mbisa, J.L. et al. Human immunodeficiency virus type 1 cDNAs
699 produced in the presence of APOBEC3G exhibit defects in plus-strand
700 DNA transfer and integration. *J Virol* **81**, 7099-110 (2007).
- 701 28. Newman, E.N. et al. Antiviral function of APOBEC3G can be
702 dissociated from cytidine deaminase activity. *Curr Biol* **15**, 166-70
703 (2005).
- 704 29. Mbisa, J.L., Bu, W. & Pathak, V.K. APOBEC3F and APOBEC3G inhibit
705 HIV-1 DNA integration by different mechanisms. *J Virol* **84**, 5250-9
706 (2010).
- 707 30. Chaurasiya, K.R. et al. Oligomerization transforms human APOBEC3G
708 from an efficient enzyme to a slowly dissociating nucleic acid-binding
709 protein. *Nat Chem* **6**, 28-33 (2014).
- 710 31. Wang, X. et al. The cellular antiviral protein APOBEC3G interacts with
711 HIV-1 reverse transcriptase and inhibits its function during viral
712 replication. *J Virol* (2012).
- 713 32. Hu, W.S. & Hughes, S.H. HIV-1 reverse transcription. *Cold Spring Harb*
714 *Perspect Med* **2**(2012).
- 715 33. Herschhorn, A. & Hizi, A. Retroviral reverse transcriptases. *Cell Mol*
716 *Life Sci* **67**, 2717-47 (2010).
- 717 34. Gillick, K. et al. The suppression of HIV-1 infection by APOBEC3
718 proteins in primary human CD4+ T cells is associated with the

719 inhibition of processive reverse transcription as well as excessive
720 cytidine deamination *submitted to J Virol* (2012).

721 35. Arts, E.J., Li, Z. & Wainberg, M.A. Analysis of Primer Extension and the
722 First Template Switch during Human Immunodeficiency Virus Reverse
723 Transcription. *J Biomed Sci* **2**, 314-321 (1995).

724 36. Driscoll, M.D., Golinelli, M.P. & Hughes, S.H. In vitro analysis of human
725 immunodeficiency virus type 1 minus-strand strong-stop DNA synthesis
726 and genomic RNA processing. *J Virol* **75**, 672-86 (2001).

727 37. Harrison, G.P., Mayo, M.S., Hunter, E. & Lever, A.M. Pausing of
728 reverse transcriptase on retroviral RNA templates is influenced by
729 secondary structures both 5' and 3' of the catalytic site. *Nucleic Acids*
730 *Res* **26**, 3433-42 (1998).

731 38. Kim, E.Y. et al. Human APOBEC3 Induced Mutation of Human
732 Immunodeficiency Virus Type-1 Contributes to Adaptation and
733 Evolution in Natural Infection. *PLoS Pathog* **10**, e1004281 (2014).

734 39. Yu, Q. et al. Single-strand specificity of APOBEC3G accounts for
735 minus-strand deamination of the HIV genome. *Nat Struct Mol Biol* **11**,
736 435-42 (2004).

737 40. Schormann, N., Ricciardi, R. & Chattopadhyay, D. Uracil-DNA
738 glycosylases-structural and functional perspectives on an essential
739 family of DNA repair enzymes. *Protein Sci* **23**, 1667-85 (2014).

740 41. Visnes, T. et al. Uracil in DNA and its processing by different DNA
741 glycosylases. *Philos Trans R Soc Lond B Biol Sci* **364**, 563-8 (2009).

742 42. Wang, Z. & Mosbaugh, D.W. Uracil-DNA glycosylase inhibitor gene of
743 bacteriophage PBS2 encodes a binding protein specific for uracil-DNA
744 glycosylase. *J Biol Chem* **264**, 1163-71 (1989).

745 43. Mansky, L.M., Preveral, S., Selig, L., Benarous, R. & Benichou, S. The
746 interaction of vpr with uracil DNA glycosylase modulates the human
747 immunodeficiency virus type 1 In vivo mutation rate. *J Virol* **74**, 7039-
748 47 (2000).

749 44. Willetts, K.E. et al. DNA repair enzyme uracil DNA glycosylase is
750 specifically incorporated into human immunodeficiency virus type 1
751 viral particles through a Vpr-independent mechanism. *J Virol* **73**, 1682-
752 8 (1999).

753 45. Adolph, M.B., Webb, J. & Chelico, L. Retroviral restriction factor
754 APOBEC3G delays the initiation of DNA synthesis by HIV-1 reverse
755 transcriptase. *PLoS One* **8**, e64196 (2013).

756 46. Voss, T.C., Demarco, I.A. & Day, R.N. Quantitative imaging of protein
757 interactions in the cell nucleus. *Biotechniques* **38**, 413-24 (2005).

758 47. Burnett, A. & Spearman, P. APOBEC3G multimers are recruited to the
759 plasma membrane for packaging into human immunodeficiency virus
760 type 1 virus-like particles in an RNA-dependent process requiring the
761 NC basic linker. *J Virol* **81**, 5000-13 (2007).

762 48. Wiegand, H.L., Doehle, B.P., Bogerd, H.P. & Cullen, B.R. A second
763 human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1
764 and HIV-2 Vif proteins. *EMBO J* **23**, 2451-8 (2004).

765 49. Wedekind, J.E. et al. Nanostructures of APOBEC3G support a
766 hierarchical assembly model of high molecular mass ribonucleoprotein
767 particles from dimeric subunits. *J Biol Chem* **281**, 38122-6 (2006).

- 768 50. Gallois-Montbrun, S. et al. Comparison of cellular ribonucleoprotein
769 complexes associated with the APOBEC3F and APOBEC3G antiviral
770 proteins. *J Virol* **82**, 5636-42 (2008).
- 771 51. Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F. & Malim,
772 M.H. RNA-dependent oligomerization of APOBEC3G is required for
773 restriction of HIV-1. *PLoS Pathog* **5**, e1000330 (2009).
- 774 52. Xiao, X., Li, S.X., Yang, H. & Chen, X.S. Crystal structures of
775 APOBEC3G N-domain alone and its complex with DNA. *Nat Commun*
776 **7**, 12193 (2016).
- 777 53. Menendez-Arias, L., Sebastian-Martin, A. & Alvarez, M. Viral reverse
778 transcriptases. *Virus Res* (2016).
- 779 54. Levin, J.G., Mitra, M., Mascarenhas, A. & Musier-Forsyth, K. Role of
780 HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol* **7**,
781 754-74 (2010).
- 782 55. Basu, V.P. et al. Strand transfer events during HIV-1 reverse
783 transcription. *Virus Res* **134**, 19-38 (2008).
- 784 56. Iwatani, Y., Rosen, A.E., Guo, J., Musier-Forsyth, K. & Levin, J.G.
785 Efficient initiation of HIV-1 reverse transcription in vitro. Requirement
786 for RNA sequences downstream of the primer binding site abrogated
787 by nucleocapsid protein-dependent primer-template interactions. *J Biol*
788 *Chem* **278**, 14185-95 (2003).
- 789 57. Masuda, T. et al. Fate of HIV-1 cDNA intermediates during reverse
790 transcription is dictated by transcription initiation site of virus genomic
791 RNA. *Sci Rep* **5**, 17680 (2015).
- 792 58. Kharytonchyk, S. et al. Transcriptional start site heterogeneity
793 modulates the structure and function of the HIV-1 genome. *Proc Natl*
794 *Acad Sci U S A* **113**, 13378-13383 (2016).
- 795 59. Klarmann, G.J., Schaubert, C.A. & Preston, B.D. Template-directed
796 pausing of DNA synthesis by HIV-1 reverse transcriptase during
797 polymerization of HIV-1 sequences in vitro. *J Biol Chem* **268**, 9793-802
798 (1993).
- 799 60. Altfeld, M. & Gale, M., Jr. Innate immunity against HIV-1 infection. *Nat*
800 *Immunol* **16**, 554-62 (2015).
- 801 61. Ahn, J. et al. HIV-1 Vpr loads uracil DNA glycosylase-2 onto DCAF1, a
802 substrate recognition subunit of a cullin 4A-ring E3 ubiquitin ligase for
803 proteasome-dependent degradation. *J Biol Chem* **285**, 37333-41
804 (2010).
- 805 62. Steagall, W.K., Robek, M.D., Perry, S.T., Fuller, F.J. & Payne, S.L.
806 Incorporation of uracil into viral DNA correlates with reduced replication
807 of EIAV in macrophages. *Virology* **210**, 302-13 (1995).
- 808 63. Kennedy, E.M. et al. Abundant non-canonical dUTP found in primary
809 human macrophages drives its frequent incorporation by HIV-1 reverse
810 transcriptase. *J Biol Chem* **286**, 25047-55 (2011).
- 811 64. Yan, N., O'Day, E., Wheeler, L.A., Engelman, A. & Lieberman, J. HIV
812 DNA is heavily uracilated, which protects it from autointegration. *Proc*
813 *Natl Acad Sci U S A* **108**, 9244-9 (2011).
- 814 65. Hansen, E.C. et al. Diverse fates of uracilated HIV-1 DNA during
815 infection of myeloid lineage cells. *Elife* **5**(2016).

- 816 66. Weil, A.F. et al. Uracil DNA glycosylase initiates degradation of HIV-1
817 cDNA containing misincorporated dUTP and prevents viral integration.
818 *Proc Natl Acad Sci U S A* **110**, E448-57 (2013).
- 819 67. Chen, R., Wang, H. & Mansky, L.M. Roles of uracil-DNA glycosylase
820 and dUTPase in virus replication. *J Gen Virol* **83**, 2339-45 (2002).
- 821 68. Sire, J., Querat, G., Esnault, C. & Priet, S. Uracil within DNA: an actor
822 of antiviral immunity. *Retrovirology* **5**, 45 (2008).
- 823 69. Stremlau, M. et al. The cytoplasmic body component TRIM5alpha
824 restricts HIV-1 infection in Old World monkeys. *Nature* **427**, 848-53
825 (2004).
- 826 70. Stremlau, M. et al. Specific recognition and accelerated uncoating of
827 retroviral capsids by the TRIM5alpha restriction factor. *Proc Natl Acad*
828 *Sci U S A* **103**, 5514-9 (2006).
- 829 71. Malim, M.H. & Bieniasz, P.D. HIV Restriction Factors and Mechanisms
830 of Evasion. *Cold Spring Harb Perspect Med* **2**, a006940 (2012).
- 831 72. Warren, K., Warrilow, D., Meredith, L. & Harrich, D. Reverse
832 Transcriptase and Cellular Factors: Regulators of HIV-1 Reverse
833 Transcription. *Viruses* **1**, 873-94 (2009).
- 834 73. Salter, J.D., Bennett, R.P. & Smith, H.C. The APOBEC Protein Family:
835 United by Structure, Divergent in Function. *Trends Biochem Sci* **41**,
836 578-94 (2016).
- 837

Online Methods

Plasmid constructs

The expression vector for Vif-deficient HIV-1_{NL4.3} has been described previously¹. Wild type and mutant A3G proteins carrying the carboxy-terminal 3xHA tag were expressed using pCMV4_HA^{2,3}. Site-directed mutagenesis was carried out based on the QuikChange protocol (Stratagene) using Pfu polymerase (Stratagene). Transfer plasmids for the generation of recombinant baculovirus were generated as follows: wild type or mutant A3G sequences were PCR amplified and cloned into pCGTHCF_{FL}T7 with a carboxy-terminal 6xHis tag sequence using XbaI and AgeI restriction sites. The tagged A3G insert was subsequently PCR amplified and inserted into the pVL1392 transfer vector (BD Biosciences) using EcoRI and XmaI restriction sites. Expression vectors for HIV-1 RT p51 and p66 were created as follows: a codon optimized gag-pol plasmid (pCO Gag-pol,⁴) was used as the PCR template, and fragments were subcloned into pCAGGS using XmaI/NotI (p51) or EcoRI/XmaI (p66). The reverse primers were designed with or without a 1xFLAG tag coding sequence. Expression vectors encoding truncated A3G proteins with amino-terminal GST tags were generated by PCR amplification and subcloned into pCAGGS_GST with EcoRI/XhoI. Fluorescent protein fusions were expressed from plasmid constructs carrying the respective cDNA sequence from the aforementioned vectors and subcloned in pEGFP_N1 (Clontech) and pmCherry_N1 (Clontech) vectors. The cDNA for the (human) codon optimized uracil-DNA glycosylase inhibitor (hUGI) gene was a gift from Michael Emerman⁵ and was subcloned into the MLV-based transfer vector pCMS28⁶. MLV packaging and VSV-G expression vectors have been described⁷.

Cell lines and stable cell lines

HEK293T and HeLa cells were obtained from ATCC. TZM-bl cells were obtained through the NIH AIDS Reagents Repository Program (ARRP). Cells were cultured in Dulbecco's modified Eagle's medium (Invitrogen, UK) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin.

CEM-SS cells, also from ARRP, were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium (Invitrogen, UK) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. All cells were tested and found to be negative for mycoplasma contamination. hUGI expressing HEK293T, CEM-SS and TZM-bl cells were created using standard MLV-based transduction vectors. HEK293T cells were co-transfected with expression plasmids for VSV-G, an MLV packaging plasmid and pCMS28 encoding hUGI. Transduced cells were selected with 1 µg/ml puromycin.

42

43 **Virion production**

HEK293T cells were seeded in 10 cm tissue culture plates and co-transfected with 10 µg pNL4.3/ΔVif and 0.25 to 10 µg pCMV4_A3G_HA wild type or mutant expression vectors using polyethylenimine (PEI, Molecular Biosciences). Total DNA levels were kept constant with empty pCMV4-HA vector. At 48 h, supernatants were harvested, DNase (RQ1 RNase free DNase (Promega)) treated for 1 h at 37°C and viruses then purified through a 20% (w/v) sucrose cushion at 28,000 x g for 75 min at 4°C. Viruses were resuspended in PBS, and quantified according to p24^{Gag} content using an enzyme-linked immunosorbent assay (ELISA; Perkin-Elmer).

53

54 **TZM-bl reporter assay**

TZM-bl reporter assays were utilized to measure single cycle HIV-1 infectivity. Virus inputs equivalent to 10 or 20 ng p24^{Gag} were used to challenge 0.5×10^5 cells in 48-well format. At 48 h post infection, whole-cell lysates were assayed for the induction of β-galactosidase expression using a Galacto-Star system (Applied Biosystems).

60

61 **Infections**

CEM-SS T cells were seeded at 1×10^6 cells per 0.5 ml of medium in a 12-well format. Virus equivalent to 75 ng p24^{Gag} was added and the cells were spin-infected at 2,000 x g for 2 h at 30°C. After centrifugation, cells were left to recover for 1 h in the incubator, before 2 washes with PBS to remove the inoculum. Cells were maintained in fresh media for 1 h before being harvested

67 by centrifugation, and pellets either frozen at -80°C or immediately processed
68 for DNA extraction.

69

70 **DNA extraction and quantitative real time PCR**

71 Whole cell DNA was isolated using the DNAeasy Blood and Tissue Kit
72 (Qiagen) according to the manufacturer's protocol. DNA was eluted with 200
73 µl H₂O. 17.5 µl were mixed with 2 µl Cutsmart buffer (NEB) and 0.5 µl Dnpi
74 (NEB) (to eliminate any remaining transfected plasmid) for 1 h at 37°C before
75 analysis by quantitative PCR. The remaining nucleic acid fraction was
76 reserved for library preparation. Early reverse transcription products were
77 detected using primers that amplify the region between nucleotides 500 and
78 635 of the provirus: oHC64 (5'-taactaggaacccactgc-3') and oHC65 (5'-
79 gctagagattttccacactg-3') and probe oHC66 (5'-FAM-
80 acacaacagacgggcacacacta-TAMRA-3'). Reactions were performed in
81 duplicate, in TaqMan Universal PCR master mix (no AmpErase), using 0.9
82 pmol of each primer per µl and 0.25 pmol probe per µl. After 10 min at 95°C,
83 reactions were cycled through 15 s at 95°C, followed by 1 min at 60°C for 40
84 repeats, carried out on an ABI Prism model 7900HT (Applied Biosystems).
85 pNL4.3/ΔVif, was diluted in herring sperm DNA to generate a standard curve.

86

87 **Statistical analysis**

88 Where indicated a paired or unpaired, two-tailed t-test was carried out using
89 the t-test function in GraphPad Prism® 6.0.

90

91 **MiSeq library preparation**

92 *Enrichment for HIV-1 specific sequences*

93 Whole cell DNA from infected cells was isolated as above and HIV-1 specific
94 DNA was enriched over the genomic DNA. Biotinylated oligonucleotides
95 complementary to the HIV-1 sequence (within the tRNA primer and within the
96 first 22 nt of the strong stop sequence, see exact sequences below) were
97 annealed to the cDNA sample and a pulldown with magnetic streptavidin
98 beads was performed: for each sample, 100 µl streptavidin beads (Promega,
99 Magasphere (Z5481) were washed (using a magnet) with 100 µl of bind and
100 wash buffer (BW: 5 mM Tris pH 7.5, 0.5 mM EDTA, 1 M NaCl) then blocked

101 for 10 min in casein solution at room temperature, and again washed once in
102 BW buffer. Beads were resuspended in BW buffer containing 50 pmol 5'-
103 cagtgtggaaaatctctagcag-biotin-3', 5'-cagtggcgcccgaaca-biotin-3', and 5'-biotin-
104 cagtgtggaaaatctctagcagtggcgcccgaacagggac-biotin-3'. The mixture was left
105 for 30 min at room temperature while rotating. Beads were then washed twice
106 in 1xTEN buffer (10 mM Tris HCl pH 8.0, 1 mM EDTA, 100 mM NaCl) and
107 then resuspended in 10 µl TEN buffer. This was added to 170 µl DNA and 90
108 µl 3xTEN. Samples were denatured at 92°C for 2 min and incubated at 52°C
109 for 1 h. Finally, beads were washed once with 1xTEN buffer and resuspended
110 in 50 µl H₂O. To elute the DNA, samples were heated to 92°C for 2 min, put
111 on the magnet and the supernatant removed quickly.

112

113 *Adaptor ligation*

114 The barcoded adaptor carries a 5' phosphate group (PHO) and a 3' three
115 carbon chain (C3) spacer (SpC3) on to the 3' hydroxyl group (5'-PHO-
116 tgaagagcctagtcgctgttcannnnnnctgccatagagagatcggaagagcacacgtct-SpC3-3')
117 (Integrated DNA Technologies or MWG Eurofins) and was self annealed in T4
118 DNA ligase buffer by heating to 92°C followed by slow cooling to 16°C (2%
119 slope on Eppendorf PCR machine). Ligation reactions were set up in 60 µl
120 total volume, with 6 µl T4 DNA ligase buffer, 24 µl 50% PEG-8000 (Sigma), 6
121 µl 5 M betaine (Sigma), 4 µl pre-annealed adaptor (400 pmol total), 1.2 µl T4
122 DNA ligase (NEB, 2,000,000 units/ml) and 18.8 µl DNA sample. Reactions
123 were incubated at 16°C overnight. As controls, instead of DNA samples, HTP
124 control oligos (listed in Supplemental Figure 2) at 100 pmol/µl were mixed at
125 equimolar ratios and then diluted 1 in 62,500 before being ligated as above.

126

127 *Adaptor removal and size selection*

128 Ligations were denatured by adding 30 µl gel loading buffer II (Ambion,
129 formamide buffer) and heating for 2 min at 94°C before being put on ice. 6%
130 TBE urea gels (precast, Life Technologies) were prerun in 1xTBE buffer for
131 ~15 min at 250 V, and each sample was loaded into three adjacent wells.
132 Gels were run for 20 min at 250 V before gel strips containing samples were
133 immersed in 1xTBE and SYBR Gold (Thermo Fisher). To avoid cross-
134 contamination between samples during staining and gel cutting, a maximum

135 of two samples were run per gel and gels were cut for separate staining. After
136 5 min, DNA was visualized on a visible blue light LED transilluminator (Dark
137 Reader, Clare Chemicals), gels were cut right above the adaptor and the
138 smear between the adaptor and the well was divided into three even pieces.
139 Each piece was crushed and 1 ml urea gel extraction buffer (0.5 M
140 $\text{NH}_4\text{CH}_3\text{CO}_2$, 1 mM EDTA, 0.2% SDS) was added. Tubes were rotated for 3
141 to 5 h and the suspensions were transferred into SpinX columns (0.2 μm ,
142 Acetate; Corning Costar) with a round whatman filter to prevent membrane
143 clogging. Columns were centrifuged at 14 000 x g for 1 min. The elutions were
144 combined with 3 μg polyA carrier RNA, 1 μl glycogen, and 0.7 ml isopropanol.
145 Samples were frozen overnight, centrifuged for 30 min at 14 000 x g and DNA
146 pellets were washed with 80% EtOH. Any remaining EtOH was evaporated at
147 55°C. Dried pellets were resuspended in 20 μl H_2O .

148

149 *PCR and preparation for MiSeq™ run*

150 PCR reactions contained 20 μl Accuprime Supermix I (Invitrogen), 18 μl DNA,
151 1 μl of 10 μM MP1.0+22HIV, a sequence modified multiplexing PCR primer
152 1.0 (MP1.0 (Illumina)) (5'-aatgatacggcgaccaccgagatctacactctttccctacacgacgct
153 cttccgatctcactgctagagattttccacactg-3') from MWG Eurofins and 1 μl NEB
154 index primer (NEBNext for Illumina Multiplex Oligo Kit). For each of the three
155 gel pieces from the same sample, separate PCR reactions, but the same
156 index primer was used. A 25 cycle PCR was performed at 55°C annealing
157 and 68°C extension. Half of each PCR reaction was pooled together and
158 purified with AmpureXP beads (1.8x ratio) according to the manufacturer's
159 instructions. Concentrations were determined using a Qubit Fluorometer
160 (Invitrogen) and the size range was determined by Tapestation (Agilent)
161 measurements. Typical tapestation gel images showed material to be
162 between roughly 150 bp to 500 bp long with two wide peaks at about 160 bp
163 and 350 bp. On initial library preparation optimization, TopoTA cloning and
164 individual sequencing of 10 or more colonies was performed to check inserts.
165 The final library concentration was adjusted to 4 nM.

166

167 **MiSeq™ library runs and data analysis**

168 Libraries were run on an Illumina MiSeq™ Benchtop Sequencer using the
169 MiSeq™ Reagent Kit v3 (150 cycle) and the 'Generate FASTQ' workflow at
170 University College London (UCL) Institute for Neurology Next Generation
171 Sequencing Facility or at the King's College London Genomics Centre. Read
172 lengths were set at 26 bases for Read1 and 142 or 125 bases for Read2. The
173 two libraries from which data are shown in this article had passfilter
174 percentages (PF%) of 93.68% and 93.11% with quality scores (AVG % Q30)
175 of 89.05% and 82.45%, respectively.

176

177 The main information extracted from each read were: first, the last nucleotide
178 of the HIV-1 sequence adjacent to the fixed adaptor sequence, which was
179 ligated to the viral cDNAs (see Figure 1c and Supplementary Figure 1b). This
180 represents the open 3'-terminus of the viral cDNA at time of harvest and,
181 second, the base variation of all bases, in particular C to T mutations. For this
182 purpose FASTQ files were subjected to in house analysis. Adaptors were
183 trimmed and sequences that were duplicated (including the barcode) were
184 removed as PCR artifacts. The remaining sequences were aligned to the HIV-
185 1 sequence using Bowtie (<http://bowtie-bio.sourceforge.net/index.shtml>),
186 allowing a maximum of 3 base mismatches, and the position of 3'-termini for
187 each read was determined from the alignment position. Mutation rates from
188 the template sequence for each base were also calculated. Where required, a
189 linear length-dependent correction factor was calculated from synthesized
190 oligos control library (see Supplementary Figure 2) and applied to the dataset
191 to correct for differences in sequencing efficiency of longer products.

192

193 Output files for each indexed sample were designed to yield the number of
194 total unique reads for each nucleotide position in the (-)sss product of the HIV-
195 1_{NL4.3} proviral sequence. The template that was used spanned the strong stop
196 and first strand transfer sequences up to the polypurine tract (U5-R-U3-PPT;
197 635 bases total). Of note, only sequences of a minimal length of 22 nt are
198 accounted for in this library due to primer design. Accordingly, there were 613
199 possible termini for the cDNA sequences in total, though our main analysis
200 focuses on the 160 nt up to base 182, the strong stop site. For base
201 substitution analysis, parse results presented the number of each of the four

possible base calls for each nucleotide position. Percentages could therefore be calculated from the total coverage of each base. Notably, coverage for each base itself is not constant for two reasons: one, the 3'-termini of individual reads vary, thereby coverage ends at varying sites; and, two, Read2 sequence length of 142 or 125 (depending on the library) may not reach the 5' end of strong stop, depending on the starting position.

Relative abundance of cDNA along the (-)sss sequence in the main figures was calculated by dividing the number of total reads for each nt position by the number of total reads up to nt 182 (for total reads see Supplementary Fig 4b). The sole exception is Supplementary Figure 4a, which shows profiles beyond first strand transfer, where the read number was divided by the total read count in the entire sample. All figures displaying cDNA profiles (Fig 1g, 2d, 6e, and Supplementary Fig 2a, 4a, 5 and 7) show the relative abundance of HIV-1 cDNA molecules for each length between nt positions 23 and 182 of the HIV-1_{NL4.3} (-)sss product (in blue histogram bars, scale on the left y-axis). All positions with cytosine bases in the HIV-1_{NL4.3} (-)sss sequence were analyzed for the presence of cytosine versus thymine/uracil bases as described above; shown in dashed red lines is the percentage of reads, which carried C to T/U mutations at the indicated position (scale on the right y-axis). Labels to the right of the graphs describe the virions used for infection.

Data and code availability

The data supporting this study and custom software are available from the corresponding author upon reasonable request. There are no restrictions to data availability. Raw MiSeq® sequencing files analyzed in this study (presented in Fig 1g, 2d, 6e as well as Supplementary Fig 2, 4, 5 and 7) are also publicly available at the European Nucleotide Archive (ENA) under study accession number PRJEB22170. Individual accession codes for each sample are listed in Supplementary Figure 4b. Custom computer code used to analyze the raw MiSeq® reads is deposited in GitHub and publicly available (doi:10.5281/zenodo.1004571).

Co-immunoprecipitation assays

236 HEK293T cells were transfected in 6 well format using PEI. At 36 h, cells were
237 scraped in ice-cold PBS, pelleted (2 min, 500 x g) and lysed in DMEM (Gibco)
238 + 0.5% IGEPAL CA 630 (Sigma) plus protease inhibitor cocktail (Roche).
239 After 10 min on ice, cell lysates were sonicated for 10 s, and clarified by
240 centrifugation at 1000 x g for 10 min. Magnetic protein G Dynabeads
241 (Invitrogen) were washed twice in PBS using a magnetic stand and then pre-
242 incubated with FLAG_M2 antibody (Sigma). The charged magnetic beads
243 were washed once, distributed to the different cell lysates and rocked at 4°C
244 for 2 h. Beads were washed four times with DMEM + 0.5% IGEPAL CA 630
245 and proteins then eluted with protein loading buffer (according to Laemmli),
246 resolved by SDS PAGE and analyzed by standard immunoblot analysis.
247 Antibodies used include anti-HA-HRP (3F10, Roche), unconjugated anti-HA
248 (3F10, Roche,)anti-FLAG M2-HRP (Sigma), anti-RT (mAb21, NIH AIDS
249 Reagents Repository Program (ARRP)), anti-p24^{Gag} (mouse monoclonal, 24-
250 2⁸) and anti-GST-HRP (RPN1236, Sigma). Horseradish peroxidase-
251 conjugated primary or secondary antibodies were visualized by enhanced
252 chemiluminescence (Pierce) and unconjugated primary antibodies were
253 detected by infrared IRDye-conjugated secondary antibodies (LI-COR
254 Biosciences) and LI-COR infrared imaging technology (LI-COR UK Ltd.) All
255 blots in their un-cropped versions and including molecular weight markers are
256 presented in Supplementary Fig 11.

257

258 For co-immunoprecipitation experiments with FLAG_RT and GST_A3G
259 truncations, 50 mM NaCl was added to DMEM + 0.5% IGEPAL CA 630 during
260 washes for more stringency. Additionally, proteins were eluted after the 2 h
261 incubation by competition with 3xFLAG peptide (150 µg/ml) for 30 min instead
262 of direct elution by protein loading buffer.

263

264 For co-immunoprecipitation coupled with RNase treatment, all starting
265 materials were doubled and the reactions were carried out as above until the
266 2 h incubation. Magnetic beads were washed twice in DMEM + 0.5% IGEPAL
267 CA 630 and reactions were split in three parts with one part left untreated and
268 RNaseA (Sigma) or RNase Mix (Roche) added to the two remaining reactions
269 at concentrations up to 100 µg/ml or 20 µg/ml, respectively. All reactions were

270 incubated at room temperature for 30 min while rocking and then washed two
271 more times in DMEM + 0.5% IGEPAL CA 630 before elution in protein loading
272 buffer.

273

274 **Surface plasmon resonance**

275 *Purification of p51 for surface plasmon resonance experiments*

276 HIV-1 p51 RT was purified from 20 10 cm dishes of HEK293T cells
277 transfected with p51_FLAG using anti-FLAG magnetic affinity resin (Sigma).
278 Briefly, at 48 h after transfection, cells were lysed in 50 mM Tris HCL 7.6, 150
279 mM NaCl, 0.5% Triton, 1 mM EDTA supplemented with 50 µg/ml RNase A
280 (Sigma) and protease inhibitor (Roche), and incubated for 10 min on ice.
281 Lysates were sonicated, clarified by centrifugation, and incubated with anti-
282 FLAG M2 magnetic beads (Sigma) for 3 h. The beads were washed
283 extensively with lysis buffer and eluted three times with lysis buffer containing
284 150 µg/ml 3xFLAG peptide (Sigma). Elutions were combined and dialyzed
285 against PBS + 0.01% IGEPAL CA 630 (Sigma) to remove the peptides.
286 Samples were concentrated to 10 µM as determined by Bradford assay, and
287 homogeneity was judged by Coomassie staining (see Supplemental Figure 8).
288

289 *Surface plasmon resonance experiments*

290 p51_FLAG was diluted to 8 µM before preparing a 2-fold dilution series down
291 to 0.16 µM. A3G was purified as below, dialyzed against PBS + 0.01%
292 IGEPAL CA 630 and used at a final concentration of 10 nM. Interaction
293 analyses were performed on a Biacore T200 instrument (GE Healthcare).
294 Binding surfaces for 6xHis tagged A3G were created by immobilizing an anti-
295 His antibody (Biacore/GE Healthcare) onto a Series S CM5 sensor chip
296 surface (Biacore/GE Healthcare) by amine coupling. The capture of the A3G
297 ligand was carried out at low surface density (~200 RU) to minimize potential
298 A3G-A3G interactions and ensure monomeric interactions with p51. The p51
299 analyte was injected at different concentrations, each in duplicate, in running
300 buffer (PBS + 0.01% IGEPAL CA 630) at 15 µl/min for 4 min followed by a 15
301 min dissociation time. Standard double referencing data subtraction methods
302 were applied⁹.

303

304 **Fluorescence resonance energy transfer (FRET) detected by**
305 **fluorescence lifetime imaging microscopy (FLIM)**

306 The interaction between A3G and HIV-1 RT subunits was measured by
307 measuring FRET between fluorescent protein tags on the respective proteins
308 of interest. Energy transfer between fluorescent proteins quickly loses its
309 efficiency as it decreases to the sixth power of distance and is limited to a
310 distance of less than 10 nm. The detection of FRET suggests proximity of two
311 fluorophores on the scale of Angstrom¹⁰. The efficiency of FRET was
312 determined by FLIM where fluorescent lifetime refers to the period of time a
313 fluorescent molecule stays in an excited state and emitting a photon. FRET
314 leads to a decrease in the fluorescence lifetime of the donor molecule that can
315 be very accurately measured using time-correlated single photon counting.
316 Contrary to other spectral methods of measuring FRET, such as sensitized
317 emission FRET (SE-FRET or ratiometric imaging) or acceptor photobleaching,
318 the ability to measure the fluorescence lifetime of fluorescent proteins
319 expressed in cells is independent of relative probe concentrations and
320 intensities, as well as being independent of photo-bleaching and spectral
321 bleed through¹¹⁻¹³. FLIM is therefore a sensitive and accurate approach to
322 measure FRET between two co-expressed molecules in biological specimens.

323

324 *Slide preparation*

325 HeLa cells were plated on glass coverslips and transfected using
326 Lipofectamine to express fluorescent fusion proteins. 24 h later, cells were
327 fixed with 4% paraformaldehyde, washed with PBS, quenched with 1 mg/ml
328 sodium borohydride in PBS and mounted using Mowiol (Calbiochem).

329

330 For virion samples, a drop of 5 µg/ml fibronectin in PBS was added to
331 coverslips, left for 1 h, removed and allowed to dry before addition of a
332 suspension of sucrose purified HIV-1 virions (typically 20 µl at a concentration
333 of an equivalent of 1 x 10⁷ pg p24^{Gag} protein/ml). After 1 h incubation at 37°C,
334 virions were fixed with 4% paraformaldehyde, quenched, then permeabilized
335 with 0.2% Triton before staining with labeled Fab fragments and mounting.
336 Virions were produced by co-transfection of HEK293T cells with pNL4.3/ΔVif
337 and expression vectors for fluorescent fusion proteins (A3G, CYPA or Vpr).

338

339 *Cy3 labelled Fab fragment generation*

340 To obtain RT specific Fab fragments, we first generated a polyclonal RT
341 specific antibody in rabbits, with Lampire Biological Laboratories, using full
342 length RT produced in *E. coli*. The RT expression plasmid was a kind gift from
343 Stephen Hughes and purification was performed as reported^{14,15}. Rabbit
344 serum had titers, as determined by ELISA, of 4.84×10^4 and 3.62×10^5 at day
345 30 and 50 post injection respectively. Antigen specific antibody was purified
346 from serum by first enriching IgG fractions using Melon Gel IgG purification
347 resin (Pierce) and then binding to full length RT isolated from HEK293T cells
348 and immobilized using the AminoLink Plus Immobilization Kit (Pierce). Bound
349 antibody was eluted with low pH buffer, neutralized and subsequently used for
350 Fab fragment preparation and purification using the Fab Micro Preparation Kit
351 (Pierce). The same kit was used to produce Fab fragments from the anti-
352 p24^{Gag} antibody (mouse monoclonal, 24-2⁸). All steps were carried out
353 according to the manufacturer's protocol and resulting products controlled by
354 SDS-PAGE and Coomassie staining. Lastly, Fab fragments were covalently
355 labeled with Cy3 using Cy3 Monoreactive Dye (GE Healthcare) and purified
356 from excess dye by gel filtration (PD MiniTrap G25, GE Healthcare).

357

358 *FRET-FLIM instrument*

359 FLIM was used to measure FRET between protein pairs, which allows the
360 determination of spatial protein interactions¹⁶. Time-domain FLIM was
361 performed with a multi-photon microscope system as described previously¹⁶⁻
362 ¹⁸. The system is based on a modified Bio-Rad MRC 1024MP workstation,
363 comprising a solid-state-pumped femtosecond Ti:Sapphire (Tsunami,
364 Spectra-Physics) laser system, a focal scan-head and an inverted microscope
365 (Nikon TE200). Enhanced detection of the scattered component of the
366 emitted (fluorescence) photons was afforded by the use of fast response
367 (Hamamatsu R7401-P) non-descanned detectors, developed in-house,
368 situated in the re-imaged objective pupil plane. Fluorescence lifetime imaging
369 capability was provided by time-correlated single photon counting (TCSPC)
370 electronics (Becker & Hickl, SPC 700). A 40x objective was used throughout
371 (Nikon, CFI60 Plan Fluor N.A. 1.3) and data collected at 500 ± 20 nm through

a bandpass filter (Coherent Inc. 35-5040). Images were all acquired at 256 x 256 pixel resolution. Laser power was adjusted to give average photon counting rates of the order $10^4 - 10^5$ photons s^{-1} (0.0001 to 0.001 photon counts per excitation event) to avoid pulse pile up that can lead to inaccurate lifetime quantification. Acquisition times up to 300 s at low excitation power were used to achieve sufficient photon statistics for monoexponential fitting, while avoiding either pulse pile-up or significant photobleaching. Excitation was at 890 nm. Widefield acceptor images were acquired using a CCD camera (Hamamatsu) at <100 ms exposure times.

381

382 *FRET data analysis*

383 Bulk measurements of FRET efficiency (i.e. intensity-based methods) cannot
384 distinguish between an increase in FRET efficiency (i.e. coupling efficiency)
385 and an increase in FRET population (concentration of FRET species) since
386 the two parameters are not resolved. The data presented here were analyzed
387 using a monoexponential decay model, and goodness of fit was confirmed by
388 determination of chi-squared values close to 1. The FRET efficiency is related
389 to the molecular separation of donor and acceptor and the fluorescence
390 lifetime of the interacting fraction by:

391

$$392 \quad \eta_{fret} = (R_0^6 / (R_0^6 + r^6)) = 1 - \tau_{fret} / \tau_d$$

393

394 Where R_0 is the Förster radius, r the molecular separation, η_{fret} is the lifetime
395 of the interacting fraction and τ_d the lifetime of the donor in the absence of
396 acceptor. η_{fret} and τ_d can also be taken to be the lifetime of the interacting
397 fraction and non-interacting fraction, respectively. All data were analysed
398 using TRI2 software (developed by Paul Barber, King's College London). Dot
399 plot data presented here include individual measurements as well as the
400 mean FRET efficiency from >8 cells (or >5 areas for virion FRET) per sample
401 +/- SD. Lifetime images of example cells are presented using a pseudocolour
402 scale whereby blue depicts normal GFP lifetime (no FRET) and red depicts
403 lower GFP lifetime (areas of FRET).

404

405 **Purification of A3G proteins**

406 A3G sequences were subcloned into pVL1392 transfer vector as described
407 above. Recombinant baculovirus stocks expressing wild type A3G or the
408 R24A mutant with C-terminal 6xHis tags were prepared using the BaculoGold
409 Baculovirus Expression System (BD Biosciences) on SF9 monolayer cells,
410 according to the manufacturer's protocol. High titer virus stocks were
411 achieved by 5 rounds of viral amplification on Sf9 monolayer cells at 3 days
412 each. Proteins were expressed in suspension Sf9 insect cells, cultured in
413 Sf900 II medium (Gibco) supplemented with 1% Pen/Strep (Gibco), by
414 infection with recombinant virus. Three days after infection cells were lysed in
415 50 mM Tris HCl pH 8.8, 0.5% NP40, 200 mM NaCl, 10% glycerol, 1mM DTT
416 supplemented with 50 µg/ml RNase A (Sigma) and protease inhibitor (Roche),
417 and incubated for 10 min on ice. The lysate was homogenized by two 10 s
418 sonication steps before incubation for 1 h at 25°C with slow rocking for
419 complete RNA digestion. Lysates were clarified by centrifugation at 1000 x g
420 for 10 min at 4°C, and adjusted to 0.6 M NaCl, before addition of nickel-
421 nitrilotriacetic acid-agarose (Novagen). Proteins were allowed to bind for 2 h
422 at 4°C while rotating. The suspension was centrifuged at 500 x g for 10 min at
423 4°C, the supernatant discarded, and lysis buffer added to the beads. The
424 suspension was loaded onto a poly-prep chromatography column (Bio-Rad)
425 and washed 5 times with 5 ml of wash buffer (50 mM Tris HCl pH 8.8, 125
426 mM NaCl, 10% glycerol, 1 mM DTT, 1% NP40, 50 mM imidazole) before
427 elution with 50 mM Tris HCl pH 8.8, 125 mM NaCl, 10% glycerol, 1 mM DTT,
428 1% NP40, 200 mM imidazole. To avoid protein precipitation off the column,
429 elution was performed with 1.2 ml elution buffer into 5 ml buffer (50 mM Tris
430 HCl pH 8.8, 125 mM NaCl, 10% glycerol, 1 mM DTT, 1% NP40). The solution
431 was then loaded onto a DEAE-FF sepharose column (GE Healthcare) on an
432 AKTA Purifier 10 (Pharmacia) and eluted with a gradient (0-100%) of buffer A
433 (50 mM Tris HCl pH 8.8, 0.5% NP40, 50 mM NaCl, 10% glycerol, 1 mM DTT)
434 and buffer B (50 mM Tris HCl pH 8.8, 0.5% NP40, 1 M NaCl, 10% glycerol, 1
435 mM DTT). Peak fractions were dialyzed into a suitable buffer, concentration
436 determined by Bradford assay with BSA standards, and homogeneity judged
437 by Coomassie staining (see Supplemental Figure 3 and 8).

438
439

440
441

442 **References method section:**

443

444

445

- 446 1. Phalora, P.K., Sherer, N.M., Wolinsky, S.M., Swanson, C.M. & Malim,
447 M.H. HIV-1 replication and APOBEC3 antiviral activity are not
448 regulated by P bodies. *J Virol* **86**, 11712-24 (2012).
- 449 2. Huthoff, H., Autore, F., Gallois-Montbrun, S., Fraternali, F. & Malim,
450 M.H. RNA-dependent oligomerization of APOBEC3G is required for
451 restriction of HIV-1. *PLoS Pathog* **5**, e1000330 (2009).
- 452 3. Huthoff, H. & Malim, M.H. Identification of amino acid residues in
453 APOBEC3G required for regulation by human immunodeficiency virus
454 type 1 Vif and Virion encapsidation. *J Virol* **81**, 3807-15 (2007).
- 455 4. Swanson, C.M., Sherer, N.M. & Malim, M.H. SRp40 and SRp55
456 promote the translation of unspliced human immunodeficiency virus
457 type 1 RNA. *J Virol* **84**, 6748-59 (2010).
- 458 5. Kaiser, S.M. & Emerman, M. Uracil DNA glycosylase is dispensable for
459 human immunodeficiency virus type 1 replication and does not
460 contribute to the antiviral effects of the cytidine deaminase APOBEC3G.
461 *J Virol* **80**, 875-82 (2006).
- 462 6. Gallois-Montbrun, S. et al. Antiviral protein APOBEC3G localizes to
463 ribonucleoprotein complexes found in P bodies and stress granules. *J*
464 *Virol* **81**, 2165-78 (2007).
- 465 7. Apolonia, L. et al. Promiscuous RNA binding ensures effective
466 encapsidation of APOBEC3 proteins by HIV-1. *PLoS Pathog* **11**,
467 e1004609 (2015).
- 468 8. Gaddis, N.C., Chertova, E., Sheehy, A.M., Henderson, L.E. & Malim,
469 M.H. Comprehensive investigation of the molecular defect in vif-
470 deficient human immunodeficiency virus type 1 virions. *J Virol* **77**,
471 5810-20 (2003).
- 472 9. Myszka, D.G. Improving biosensor analysis. *J Mol Recognit* **12**, 279-84
473 (1999).
- 474 10. Voss, T.C., Demarco, I.A. & Day, R.N. Quantitative imaging of protein
475 interactions in the cell nucleus. *Biotechniques* **38**, 413-24 (2005).
- 476 11. Chen, Y., Mills, J.D. & Periasamy, A. Protein localization in living cells
477 and tissues using FRET and FLIM. *Differentiation* **71**, 528-41 (2003).
- 478 12. Wallrabe, H. & Periasamy, A. Imaging protein molecules using FRET
479 and FLIM microscopy. *Curr Opin Biotechnol* **16**, 19-27 (2005).
- 480 13. Becker, W. Fluorescence lifetime imaging--techniques and
481 applications. *J Microsc* **247**, 119-36 (2012).
- 482 14. Boyer, P.L., Clark, P.K. & Hughes, S.H. HIV-1 and HIV-2 reverse
483 transcriptases: different mechanisms of resistance to nucleoside
484 reverse transcriptase inhibitors. *J Virol* **86**, 5885-94 (2012).
- 485 15. Boyer, P.L., Ding, J., Arnold, E. & Hughes, S.H. Subunit specificity of
486 mutations that confer resistance to nonnucleoside inhibitors in human
487 immunodeficiency virus type 1 reverse transcriptase. *Antimicrob*
488 *Agents Chemother* **38**, 1909-14 (1994).

- 489 16. Peter, M. et al. Multiphoton-FLIM quantification of the EGFP-mRFP1
490 FRET pair for localization of membrane receptor-kinase interactions.
491 *Biophys J* **88**, 1224-37 (2005).
492 17. Parsons, M. et al. Spatially distinct binding of Cdc42 to PAK1 and N-
493 WASP in breast carcinoma cells. *Mol Cell Biol* **25**, 1680-95 (2005).
494 18. Prag, S. et al. Activated ezrin promotes cell migration through
495 recruitment of the GEF Dbl to lipid rafts and preferential downstream
496 activation of Cdc42. *Mol Biol Cell* **18**, 2935-48 (2007).
497

Figure 1

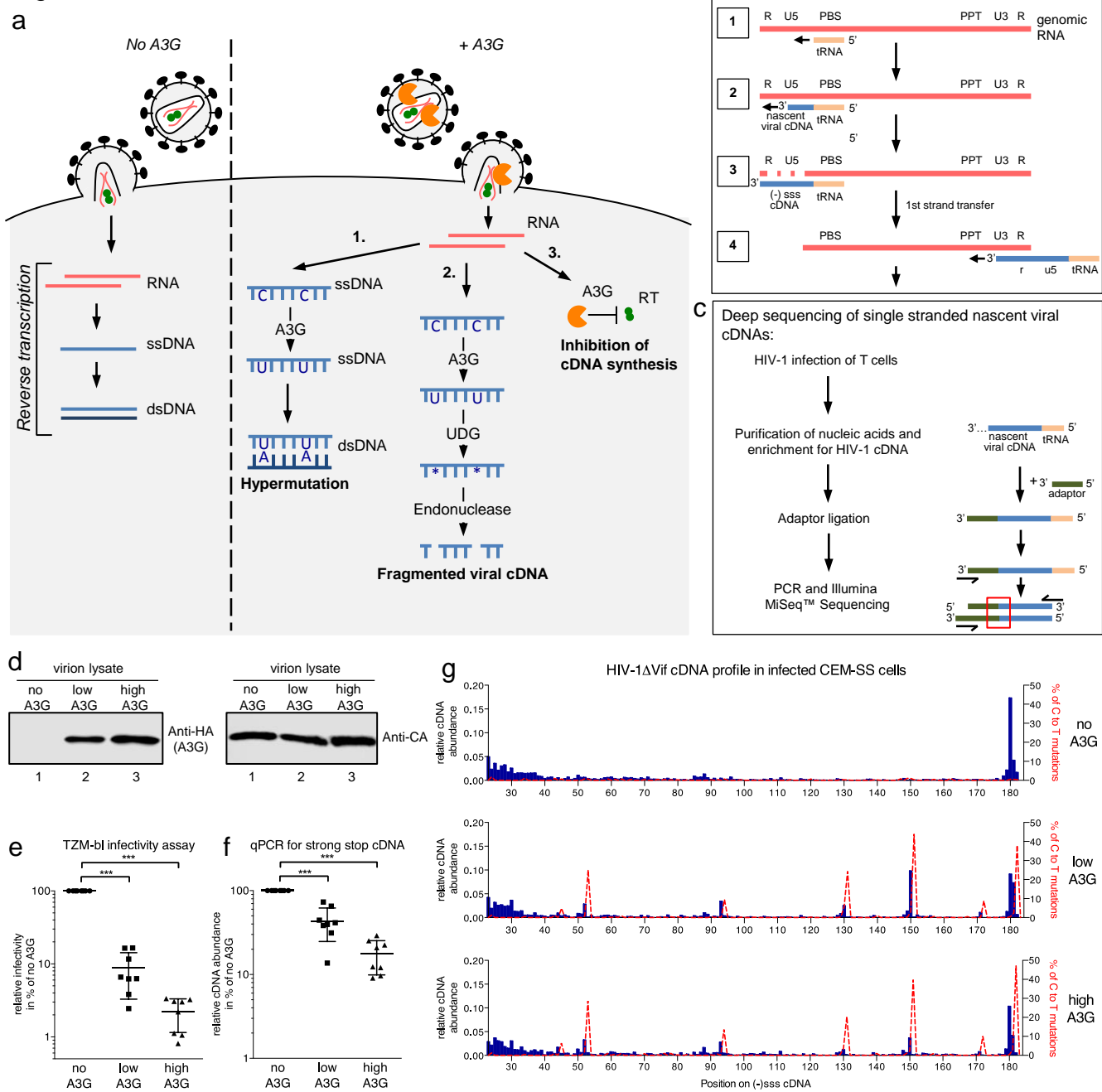


Figure 2

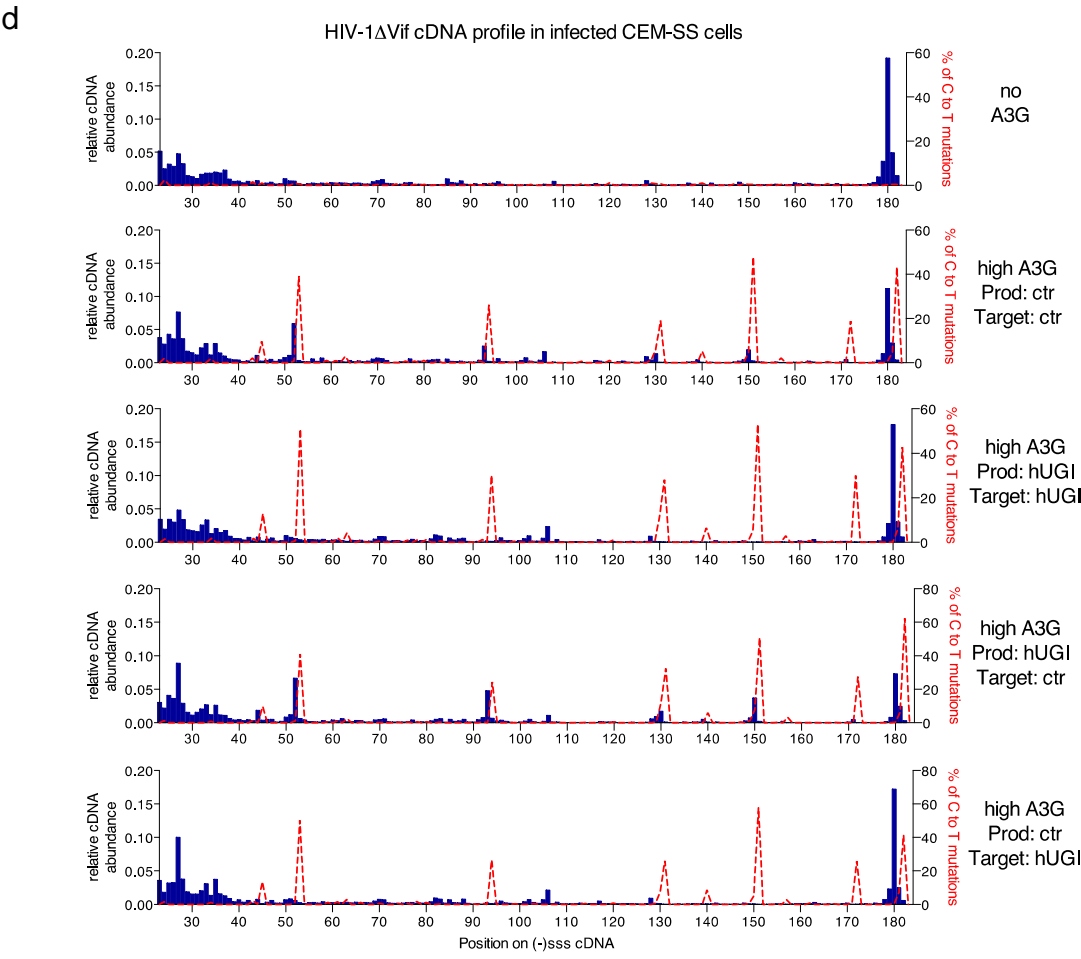
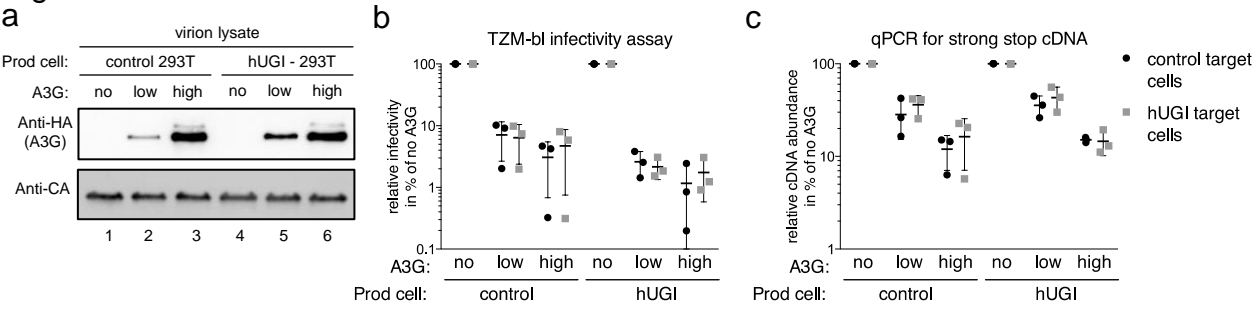


Figure 3

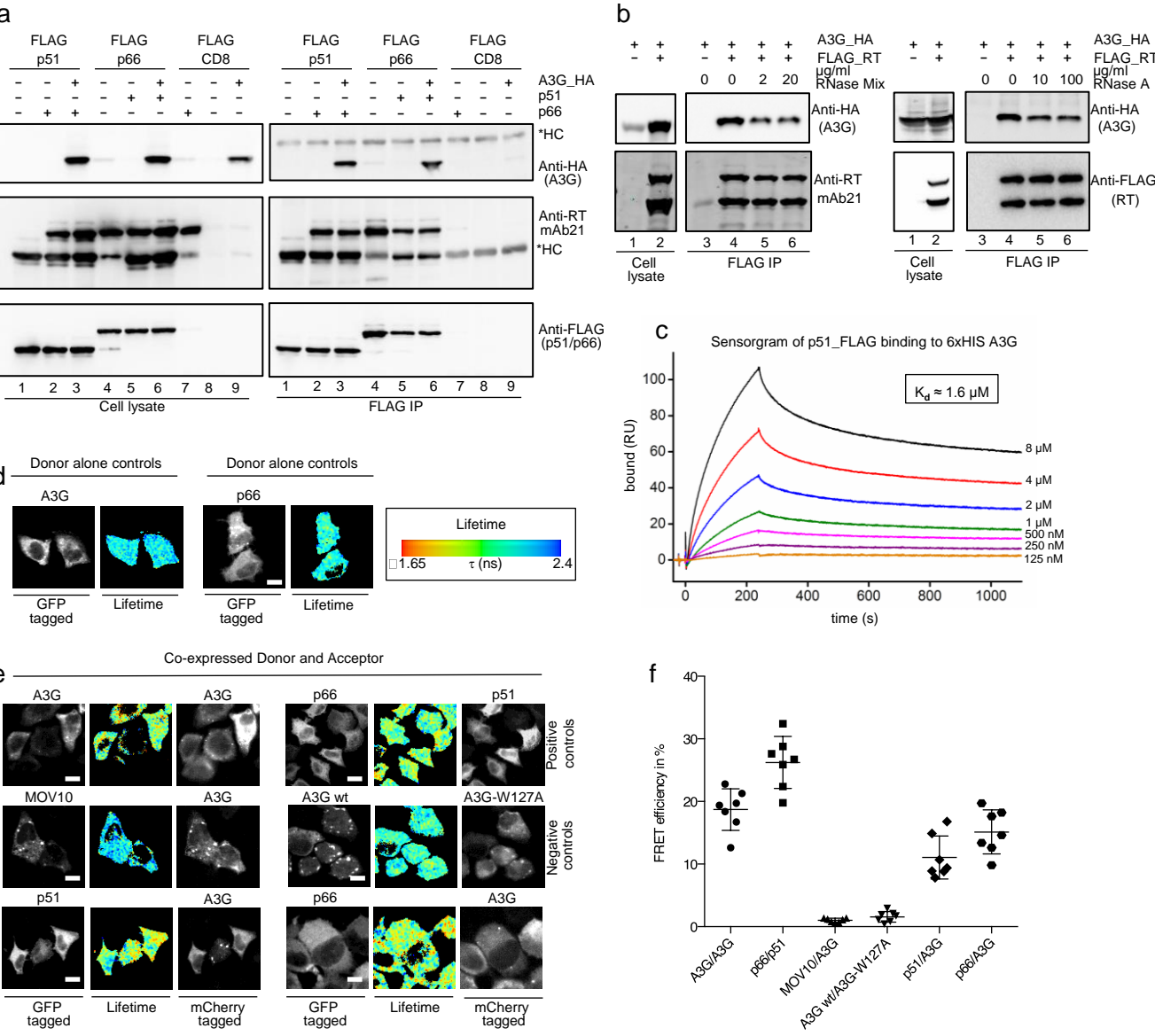


Figure 4

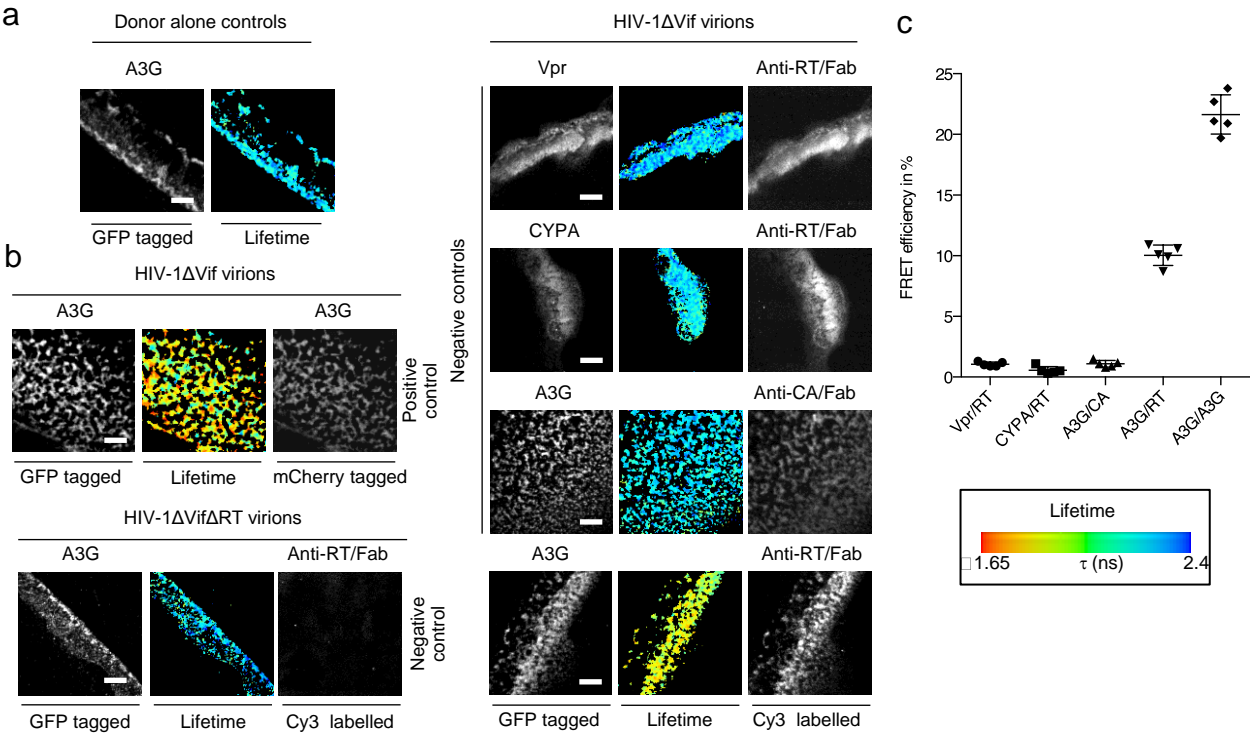


Figure 5

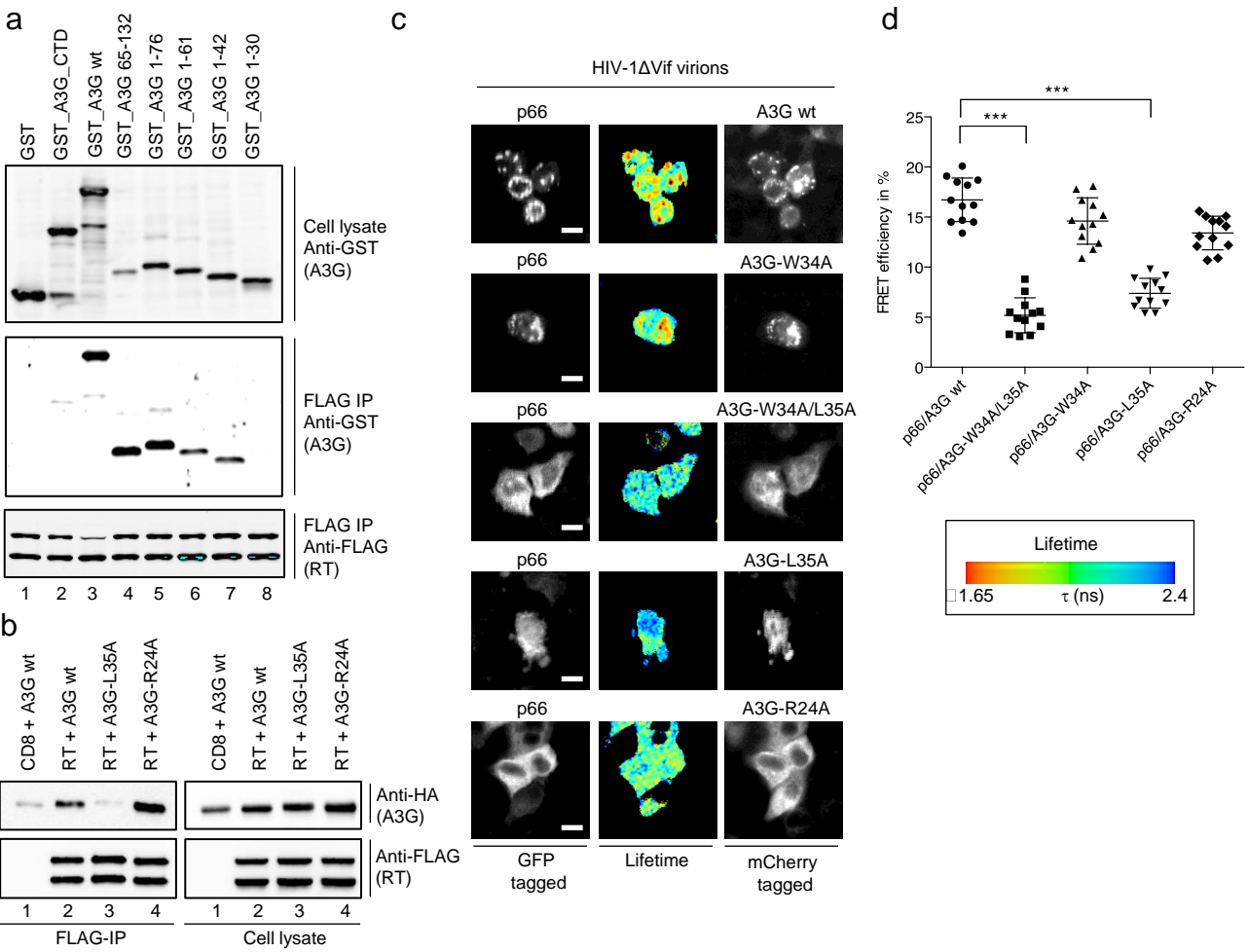


Figure 6

